NASA/TM—2003-211307

Integrating Oil Debris and Vibration Measurements for Intelligent Machine Health Monitoring

Paula J. Dempsey
Glenn Research Center, Cleveland, Ohio

March 2003

The NASA STI Program Office . . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA Scientific and Technical Information (STI) Program Office plays a key part in helping NASA maintain this important role.

The NASA STI Program Office is operated by Langley Research Center, the Lead Center for NASA's scientific and technical information. The NASA STI Program Office provides access to the NASA STI Database, the largest collection of aeronautical and space science STI in the world. The Program Office is also NASA's institutional mechanism for disseminating the results of its research and development activities. These results are published by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA's counterpart of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or cosponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services that complement the STI Program Office's diverse offerings include creating custom thesauri, building customized databases, organizing and publishing research results . . . even providing videos.

For more information about the NASA STI Program Office, see the following:

- Access the NASA STI Program Home Page at ***http://www.sti.nasa.gov***

- E-mail your question via the Internet to help@sti.nasa.gov

- Fax your question to the NASA Access Help Desk at 301–621–0134

- Telephone the NASA Access Help Desk at 301–621–0390

- Write to:
  NASA Access Help Desk
  NASA Center for AeroSpace Information
  7121 Standard Drive
  Hanover, MD 21076

NASA/TM—2003-211307

Integrating Oil Debris and Vibration Measurements for Intelligent Machine Health Monitoring

Paula J. Dempsey
Glenn Research Center, Cleveland, Ohio

National Aeronautics and
Space Administration

Glenn Research Center

March 2003

Acknowledgments

I would like to express my gratitude to the many people that have contributed to the completion of this thesis. I would like to thank my employer, NASA Glenn Research Center, more specifically James Zakrajsek, Chief of the Mechanical Components Branch. He gave me the opportunity to work in a new area and mentored me along the way. He provided advocacy and funding for this research. Without his support and guidance, this thesis would not be possible. I would like to thank my advisor, Dr. Abdollah Afjeh, and the other members of my dissertation examining committee, Dr. Kenneth DeWitt, Dr. Samuel Huang, Dr. Ahmet Kahraman, and Dr. Efstratios Nikolaidis, for their thorough review of this work. I would like to especially thank my advisor who accepted me as a student into the graduate program and signed on as my advisor. At the time, he didn't realize he accepted a challenge that would continue for almost five years. His flexibility, patience, perseverance, guidance and direction during this endeavor were truly remarkable. I would like to thank all the members of the NASA Glenn Drives Team for their assistance and support throughout this project. They were generous with their time and expertise in educating me on gear fatigue tests. They allowed me to instrument their test rigs and answered endless questions in this area. I would like to thank all the members of the NASA Glenn Technical Information Division that provided library services, photo services, and publishing expertise. I would like to thank the NASA Glenn Training Office, mainly Dennis Conrad, for providing funding for all of my graduate coursework at the University of Toledo.

Available from

NASA Center for Aerospace Information
7121 Standard Drive
Hanover, MD 21076

National Technical Information Service
5285 Port Royal Road
Springfield, VA 22100

Available electronically at http://gltrs.grc.nasa.gov

Table of Contents

Chapter 1

INTRODUCTION

1.1 Background and Motivation

One of the National Aeronautics and Space Administration's (NASA's) current goals, the National Aviation Safety Goal, is to reduce the aircraft accident rate by a factor of 5 within 10 years, and by a factor of 10 within 25 years. One of the leading factors in fatal aircraft accidents is loss of control in flight, which can occur due to flying in severe weather conditions, pilot error, and vehicle system failure. Focusing on helicopter system failures, an investigation in 1989 found that 32 percent of helicopter accidents due to fatigue failures were caused by damaged engine and transmission components (Astridge (1989)). Another report on helicopter accidents was published in July 1998 in support of the National Aviation Safety Goal (Aviation Safety and Security Program, the Helicopter Accident Analysis Team (1998)). The purpose of this study was to recommend areas most likely to reduce rotorcraft fatalities in the next ten years. A study of 1168 fatal and nonfatal accidents that occurred from 1990 to 1996 found that after human factors related causes of accidents, the next most frequent causes of accidents were due to various system and structural failures. Loss of power in-flight caused 26 percent and loss of control in-flight caused 18 percent of this type of accident. In more recent statistics, of the world total of 192 turbine helicopter accidents in 1999, 28 were directly due to mechanical failures with the most common in the drive train of the gearboxes (Learmont (2000)).

One technology area recommended for helicopter accident reduction is the design of Health and Usage Monitoring Systems (HUMS) capable of predicting imminent equipment failure for on-condition maintenance and more advanced systems capable of warning pilots of impending equipment failure. Transmission diagnostics are critical to helicopter safety and an important part of a helicopter HUMS because helicopters depend on the power train for propulsion, lift, and flight maneuvering. In order to predict transmission failures, the system must provide real-time performance monitoring of the transmission components and must also demonstrate a high level of reliability to minimize false alarms.

Various diagnostic tools exist for diagnosing damage in helicopter transmissions, the most common being vibration-based tools. Using vibration data collected from gearbox accelerometers, algorithms are developed to detect when gear damage has occurred. Over the past 25 years, numerous vibration-based algorithms for gear damage detection have been developed. Unfortunately, to this date, a complete database of existing vibration algorithms and their capabilities and limitations is not available. This is due in part to the limited transmission damage data required for assessment and validation of vibration algorithm performance.

Oil debris is another diagnostic tool used to identify abnormal wear-related conditions of transmissions. Many techniques are currently available for wear debris monitoring (Hunt (1993)). Oil debris monitoring for gearboxes consists mainly of off-line oil analysis, where samples are analyzed for trends that indicate component failure, or plug type chip detectors, where a magnet captures debris and forms an electrical bridge between contacts that indicates a state change. Inductance type sensors, used for detecting the failure of rolling element bearings in engines, but not commonly used for gear damage detection, measure a disturbance to a magnetic field caused by a particle passing through the sensor.

The goal in the development of future HUMS is to increase reliability and decrease false alarms. HUMS are not yet capable of real-time, on-line, health monitoring. Current data collected by HUMS is processed after the flight and is plagued with high false alarm rates and undetected faults. The current fault detection rate of commercially available HUMS through vibration analysis is about 70 percent (Larder (1999)). False warning rates are on average, 1 per hundred flight hours (Stewart (1997)). This is due to a variety of reasons. Vibration-based systems require extensive interpretation by trained diagnosticians. Operational effects can adversely impact the performance of vibration diagnostic parameters and result in false alarms (Dempsey and Zakrajsek (2001); Campbell, et al. (2000)). Analysis of oil debris data also requires interpretation by experts to determine the health of the monitored system. False alarms also occur when using oil debris analysis. This is due to non-failure debris, introduced into the system during routine maintenance, detected by the oil debris sensor (Howard and Reintjes (1999)).

Eurocopter, the first aircraft manufacturer to develop HUMS for its helicopters, have documented an assessment of their experience in HUMS development over the past 10 years (Pouradier and Trouve' (2001)). In this paper, they noted several shortfalls of today's HUMS, identified several reasons for these shortfalls, and offered their ideas to correct these shortfalls. For completeness, a table of these shortfalls and proposed ways of improvement has been reproduced in Table 1.1. The reasons listed, such as system complexity and damage never or inconsistently detected, confirm the need to improve the performance of current HUMS. This table also indicates the diagnostic system will be used as a maintenance tool. For this reason, the diagnostic system must be provide the end user a simple decision making tool on the health of the system.

One technique for increasing the reliability and decreasing the false alarm rate of current HUMS is to replace simple single sensor limits with multisensor systems integrating different measurement technologies. Integrating the sensors into one system is believed to be the critical key to improving damage detection. Recent papers have been published that discuss the benefits of integrating different measurement technologies such as oil and vibration based systems to improve current HUMS. One paper applied data fusion techniques to accelerometer data collected from 8 accelerometers on a helicopter gearbox (Erdley and Hall (1998)). Controlled ground tests were performed on a Chinook

TABLE 1.1
Eurocopter's list of shortfalls (Pouradier and Trouve (2001)).

Shortfalls/Unforeseen Difficulties	Reasons identified	Eurocopter's answer
Integration with operator's maintenance and logistic organization	1. System complexity 2. New operator skills	• Adaptation of organizations (done) • Training (continuing action) • Improved documentation (continuing action) • Support from aircraft manufacturer (continuing action)
Limited maintenance credits • Limited maintenance alleviation • TBOs unchanged	Performance • Lack of evidence of performance • Incomplete defect coverage • Limited prognosis performance Regulation Requirements more demanding than those for maintenance tools	Performance (launched) • Cooperation with operators on database gathering/analysis • Research activity to increase defect coverage and prognosis performance • Economic benefit of structural usage monitoring to be assessed Regulation Consider HUMS a maintenance tool
Some mechanical damage is still missed • Monitoring of epicyclic stages to be improved • Some damage is never or is inconsistently detected	Performance • Incomplete defect coverage	Performance • Research activity to increase defect coverage (continuing action) • Techniques other than vibration analysis to be considered (launched)
Operating cost higher than anticipated • Decision making sometimes difficult	Performance • Limited diagnosis perform-ance because of not "defect specific" monitoring techniques	Performance • Improved diagnostic procedures (continuing action) • Research activity to improve diagnosis performance (launched)
Acquisition cost • Most of the Civil applications in the North Sea sector • HUMS mostly installed on heavy aircraft	Technology • Not enough standardization • Difficulty in retrofitting HUMS in aircraft with analogue avionics • Rapid obsolescence Regulation • High integrity requirement	Technology • Standardization (continuing action) • Integration into digital avionics systems (done) Regulation • Consider HUMS a maintenance tool
Support cost higher than anticipated • Long maturing process • Help for diagnostics • Threshold adjustment • Continuous development	Performance • Monitoring techniques not "defect specific" Regulation • High integrity requirement	Performance (continuing action) • Streamlining ongoing development activity through support contracts • Improved diagnostic procedures • Research activity Regulation • Consider HUMS a maintenance tool

CH-46 helicopter gearbox with faults introduced into the system. The objective of this work was to classify different types of faults based on vibration data. Three decision level fusion techniques were used to make a fused decision: voting, weighted voting, and Bayesian inference. Results showed more reliable decisions were achieved through the use of multisensor fusions over the single sensor case.

Several other papers provided conceptual approaches to the fusion of oil debris and vibration measurements for condition monitoring, but did not demonstrate integration of vibration and oil debris measurement technologies results in a gear health monitoring system with improved detection and decision-making capabilities. One provided a simple framework of integrating oil and vibration technologies with a discussion on a new oil debris sensor under development by the author (Howard and Reintjes (1999)). Another paper on vibration and oil debris data collected from a small gearbox shows vibration and oil debris increase when damage occurs (Byington, et al. (1999)).

A visual programming toolkit was also developed for multisensor data fusion applications (Hall and Kasmala (1996)). The toolkit gives the user the capability to select and apply multisensor data fusion processing techniques to experimental data. Using preliminary data collected in the Spur Gear Fatigue Rig, the toolkit software was programmed. The output was a plot of darker/denser lines to indicate the possible damage. Decision-making capabilities were not part of this work.

1.2 Statement of Problem, Scope and Objectives

The basic hypothesis of this thesis is to demonstrate integrating the different measurement technologies results in a system with improved detection and decision-making capabilities as compared to existing individual diagnostic tools. Specifically, the objective of this research is to integrate oil debris and vibration based gear damage detection techniques to obtain an improved system for detecting gear pitting damage. The hypothesis will be evaluated experimentally by collecting vibration and oil debris data from fatigue tests performed in the NASA Glenn Spur Gear Fatigue Rig. The vibration data will be collected from accelerometers and used to calculate gear vibration diagnostic algorithms. The oil debris data will be collected using a commercially available in-line oil debris sensor. A gear diagnostic feature based on oil debris will also be developed as part of this thesis. Once a significant amount of experiments are performed with and without gear pitting damage, the oil debris and vibration data will be integrated using fuzzy logic and multisensor data fusion techniques combined into a system model.

Referring back to Table 1.1, this dissertation will address several shortfalls listed in this table that are also applicable to operation of the NASA Glenn fatigue test rigs. Results of this research are expected to decrease the system complexity by providing the end user with a simple tool to determine the health of a system. Moreover, addition of another measurement technology, oil debris analysis, will expand the defect coverage.

1.3 Overview of Research

The benefits of combining multiple sensors to make decisions include improved detection capabilities, decreased ambiguity, and increased probability the event is detected. However, if the sensors are inaccurate, or the features extracted from the sensors are poor predictors of transmission health, integration of these sensors will decrease the accuracy of damage prediction. For this reason, one must carefully choose the sensor and features extracted from the sensors prior to integrating data from two different measurement technologies. The reasoning behind selection of the oil debris sensor, the two vibration algorithms and the analytical approach will be discussed.

Several companies manufacture on-line inductance type oil debris sensors that measure debris size and count particles (Hunt (1993)). New oil debris sensors are also being developed that measure debris shape in addition to debris size in which the shape is used to classify the failure mechanism (Howard, et al. (1998) and Roylance (1997)). The oil debris sensor used in this analysis was selected for several reasons. The first three reasons were sensor capabilities, availability and researcher experience with this sensor. Results from preliminary research indicate that the debris mass measured by the oil debris sensor showed a significant increase when pitting damage began to occur (Dempsey (2000)). The oil debris sensor has also been used in aerospace applications for detecting bearing failures in aerospace turbine engines (Miller and Kitaljevich (2000)). From the manufacturer's experience with rolling element bearing failures, a relationship was established to set warning and alarm threshold limits for damaged bearings based on accumulated mass. Regarding its use in helicopter transmissions, a modified version of this sensor has been developed and installed in an engine nose gearbox and is currently being evaluated for an operational AH-64 (Howe and Muir (1998)). Due to limited access to oil debris data collected by this type of sensor from gear failures, no such relationship is available that defines oil debris threshold limits for damaged gears. A feature for indicating gear tooth damage and a method for defining warning and alarm limits based on this feature are required prior to data fusion. The method developed in support of this research is outlined in Chapter 3, section 3.1.4, Oil Debris Feature.

Although various techniques exist for diagnosing damage in helicopter transmissions, the method most widely used involves monitoring vibration (Land (1998)). Numerous algorithms have been developed for the processing of vibration data collected from gearbox accelerometers to detect when gear damage has occurred. Since the focus of this dissertation is not the development of new vibration algorithms for gear damage detection, the vibration-based algorithms used in this analysis were limited to those assessed in NASA Glenn test rigs. Several references summarize the transmission diagnostics using vibration-based measurement technologies tested at NASA Glenn from 1990 until 1997 (Zakrajsek et al. (1995b); Zakrajsek (1994); and Townsend (1997)). The vibration algorithms chosen for this analysis, FM4 and NA4, were selected based on their maturity, published success in detecting damage to gears, and validation in the NASA Glenn test rigs (Stewart (1977); Zakrajsek (1989); Zakrajsek (1993, 1994a, 1994b, 1995a)). FM4 was developed over 20 years ago to detect changes in the vibration pattern resulting from damage on a limited number of teeth (Stewart (1977)). NA4 was developed over 8 years ago to detect the onset of gear damage and to continue to react to the damage as it spreads (Zakrajsek, et al. (1993)). Details of both vibration algorithms can be found in Chapter 3, section 3.1.1, Vibration Features FM4 and NA4.

Prior to integrating two measurement technologies, the individual accuracy and integrity of both the oil debris sensor and the vibration algorithms must be assessed. This was done early in the research process to verify the feasibility of this research proposal.

Detailed results of these preliminary tests are discussed in Chapter 3, section 3.1.2, Preliminary Evaluation of Damage Detection Features. If during these tests, the selected oil debris sensor and vibration algorithms show no indication of damage individually, combining them will not be of much benefit. Tests on spur gears in the Spur Gear Fatigue Test Rig were conducted to establish validity of the above two measurement techniques. Additionally, experimental data from these tests were used to compare the relative performance of these methods. Results of these tests indicate the debris mass measured by the oil debris sensor is comparable to the vibration algorithms in detecting gear pitting damage. Summarized results have been published (Dempsey (2000)). From these results it was determined conclusively that the research objectives were feasible and successful results will have the potential to improve the design of future HUMS (Forror (2000)).

This research focused on one type of gear and one mode of gear damage, spur gears and pitting damage. Spur gears were chosen based on the availability of aerospace quality test gears in the Spur Gear Fatigue Test Rig. Pitting is a fatigue failure due to the high contact stresses found in gears. Pitting occurs when small pieces of material break off from the gear surface, producing pits on the contacting surfaces (Townsend (1991)). Pitting fatigue was chosen as the failure mechanism because of the availability of pitting fatigue damage data in the Spur Gear Fatigue Test Rig and because NASA's goal is to design safer drive trains through the design of gears that do not fail catastrophically without warning. Since fatigue cracks, in many cases, propagate quickly, design guidelines have been established to prevent these catastrophic failures (Lewicki (2001)). Future gears will be designed that fail in the most benign and detectable manner.

Gears were run until pitting occurs on several teeth. Pitting was detected by visual observation through periodic inspections on the first two experiments performed with damage. Pitting was detected by a video inspection system on the remaining experiments with pitting damage. The video inspection system installed on the rig is capable of following the progression of gear pitting without gearbox cover removal. Two levels of pitting were monitored per standard Spur Gear Fatigue Test Procedures, initial and destructive pitting. In this study, initial pitting is defined as pits less than 1/64 in. (0.0397cm) diameter and cover less than 25 percent of tooth contact area and destructive pitting is more severe and defined as pits greater than 1/64 in. (0.0397cm) diameter and cover greater than 25 percent of tooth contact area. If not detected in time, destructive pitting can lead to a catastrophic transmission failure if the gear teeth crack.

Multisensor data fusion analysis techniques were chosen for application to gear damage data collected from two accelerometers and an oil debris sensor in the NASA Glenn Spur Gear Fatigue Test Rig. Multisensor data fusion is a process similar to methods humans use to integrate data from multiple sources and senses to make decisions. In this process, data from multiple sensors are combined to perform inferences that are not possible from a single sensor. Commercially available software, Matlab®, was used to perform the analysis on the data collected in support of this thesis.

This thesis is organized as follows. Chapter 2 describes the NASA Glenn Spur Gear Fatigue Test Rig, test procedures, instrumentation, and data collection. Chapter 3 discusses the research methodology and is separated into 3 subchapters. Subchapter 3.1 outlines the diagnostic feature selection and validation in four sections. Section 3.1.1 discusses feature extraction from vibration data using algorithms FM4 and NA4. Feature extraction refers to the process that converts the data output from the sensor into a representation of the data that is useful in the damage identification process. Section 3.1.2 discusses preliminary data used to assess the feasibility of this topic. Section 3.1.3 discusses vibration algorithm NA4 Reset developed as the result of the preliminary evaluation of the vibration data discussed in section 3.1.2. Section 3.1.4 defines the process used to develop the oil debris damage detection feature. Subchapter 3.2 provides an overview of the data analysis methods in two sections. Section 3.2.1outlines the data fusion process. Section 3.2.2 discusses fuzzy logic, the analysis technique chosen for the damage identification and decision fusion steps in the data fusion process. Subchapter 3.3 discusses the validation of the features for damage detection. Chapter 4 discusses the results of this research in two subchapters. Subchapter 4.1 provides an assessment of the integration of the diagnostic features. Subchapter 4.2 discusses applying this analysis to the NASA Glenn Spiral Bevel Gear Test Facility. Conclusions and future work are presented in Chapter 5. The provided appendices contain supplementary analyses and discussions in support of various topics and are referred to in the respective chapters.

Chapter 2

EXPERIMENTAL SETUP AND PROCEDURES

Experimental data for this research were obtained from tests performed in the Spur Gear Fatigue Test Rig at NASA Glenn Research Center. The Spur Gear Fatigue Test Rig became operational in 1972. The rig was developed to study the effects of gear materials, gear surface treatments and lubrication on the surface fatigue strength of aircraft quality gears. The fatigue rig was modified later for use in diagnostic studies. Diagnostic tests began to be performed in conjunction with the fatigue tests in 1992 (Zakrajsek, et al. (1992)). The fatigue rig is capable of loading gears, then running gears until pitting failure is detected. Figure 2.1 shows the test apparatus in a schematic drawing. Figure 2.2 shows the test apparatus in a cutaway view. Operating on a four square principle, the shaft and gear on one shaft are coupled together with a torque applied by a hydraulic loading mechanism that twists the shafts with respect to one another. The power required to drive the system is only enough to overcome friction losses in the system (Lynwander (1983)).

The test gears are standard spur gears having 28 teeth, 3.50 in. (8.89 cm) pitch diameter, and 0.25 in. (0.635 cm) face width. The test gears were made from SAE 9310 or Pyrowear 53 steel and manufactured to AGMA class 13 aircraft tolerances. The test gears are run offset, as shown in Figure 2.2, to provide a narrow effective face width to maximize gear contact stress while maintaining an acceptable bending stress. Offset testing also allows four tests on one pair of gears.

Fatigue test procedures allowed damage to be correlated to the oil debris and vibration sensor data and are discussed in the following section. For these tests, the shaft speed was 10000 RPM and applied torque was either 53 or 71 ft-lbs (72 or 96 N·m). The load was increased during some tests to obtain failures within a shorter test time. Prior to collecting test data, the gears were subjected to break in operation for 1 hr at a torque of 10 ft-lbs (14 N·m) and 10000 RPM. Test gears were inspected periodically for damage throughout the duration of the test. The data measured during this break-in period was stored, then the oil debris sensor was reset to zero at the start of the loaded test. Test gears were inspected either manually or using a video inspection system. The video inspection system consisted of a micro camera, VCR, and monitor. Use of the video inspection system did not require gearbox cover removal. The micro camera was fed through 2 ports on the top of the gearbox cover. One port is shown in Figure 2.1 as the viewing port. The contact surface area image was then recorded for each tooth. When damage was found, the damage was documented and correlated to the test data based on a reading number. In order to document tooth damage, reference marks were made on the driver and driven gears during installation to identify tooth 1. The mating teeth numbers on the driver and

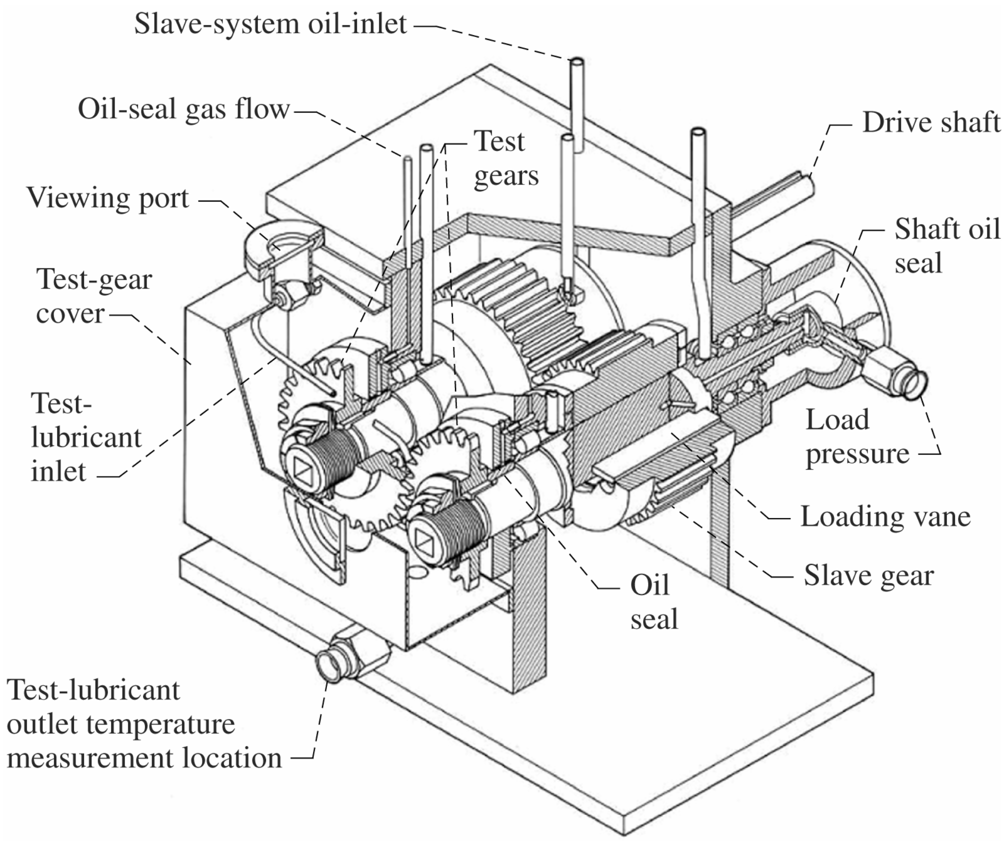


Figure 2.1.—Spur gear fatigue test rig.

driven gears are then numbered from this reference. Figure 2.3 identifies the driver and driven gear with the gearbox cover removed.

Data were collected using vibration, oil debris, speed and pressure sensors installed on the test rig. Detailed sensor specifications are listed in Table 2.1.

Vibration was measured on the gear housing and through the shaft using miniature, lightweight, piezoelectric accelerometers. Locations of both sensors, labeled shaft and housing, are shown in Figure 2.3. These locations were chosen based on an analysis of optimum accelerometer locations for this test rig (Zakrajsek, et al. (1992)). A modal analysis was also performed on the rig to verify these locations did not change. Results of this analysis are in Appendix C. The two sensors were chosen based on their ability to measure a high frequency signal, since the gear meshing frequency for 28 teeth is approximately 4700 Hz (28 teeth by 10,000 rpm by 1/60), and the first harmonic is 9400 Hz. The vibration data were sampled at 200 KHz. Per the Nyquist theorem, the sample rate must be 2 times the maximum frequency component in the signal measured. This is a starting point for adequate sampling rate, with the actual sampling rate at 5 or 10 times the maximum frequency. In some cases, the sampling rate is limited by the data acquisition (DAQ) card. The DAQ card used for this application is capable of sampling 8 differential channels at a net sampling rate of 1250 KHz. Since only four channels were recorded on the DAQ card, sample rates of 200 KHz could be obtained. In order to verify

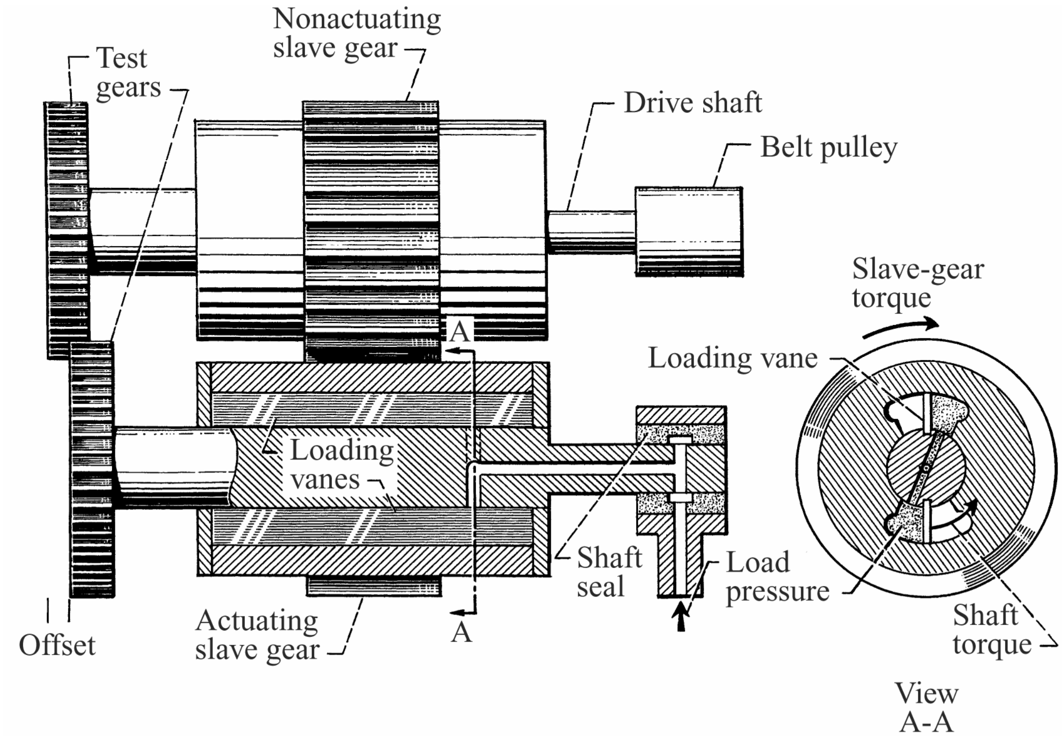


Figure 2.2.—Spur gear fatigue rig cutaway view.

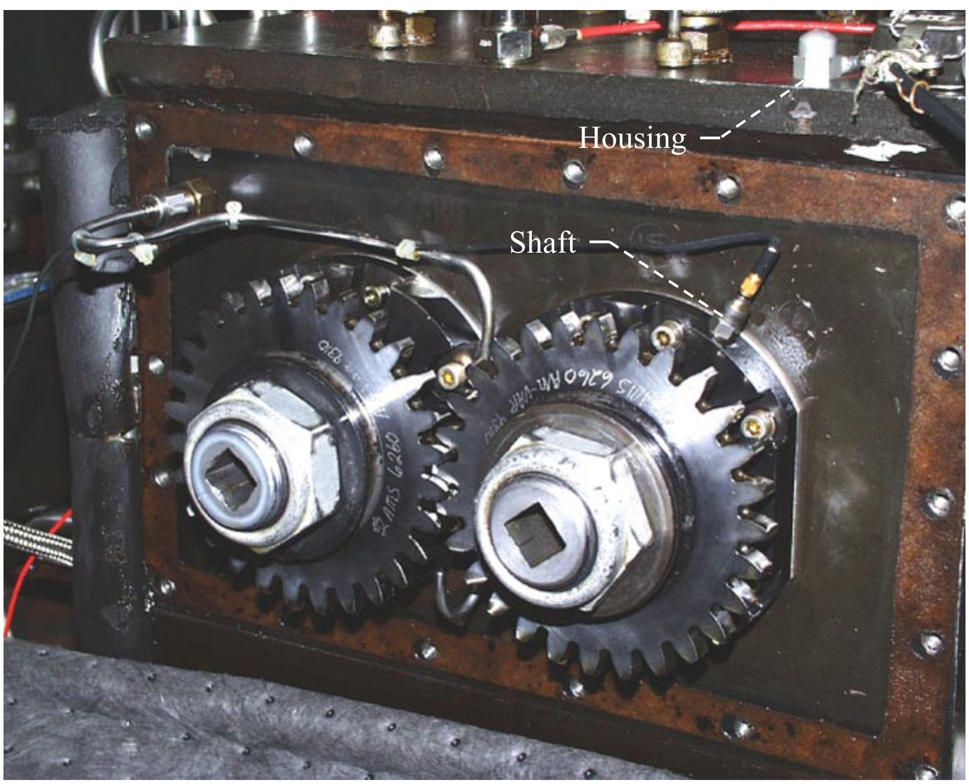


Figure 2.3.—Accelerometer locations on spur gear fatigue test rig (gearbox cover removed).

TABLE 2.1
Sensor Specifications

Sensor	Range
Shaft Accelerometer	0.7 Hz - 10 KHz
Housing Accelerometer	5 Hz - 30 KHz
Oil Debris	125-1016 microns
Shaft Speed	0-10,000 RPM
Load Pressure	0-1000 psi

frequencies over half the sampling frequency were not recorded, a lowpass filter was added before the vibration signal entered the DAQ card.

Oil debris data were collected using a commercially available oil debris sensor. The oil debris sensor is an in-line device installed downstream of the test lubricant outlet identified on Figure 2.1. The sensor consists of three coils surrounding a nonconductive section of tubing. Two coils are wound in opposite directions and are driven by an alternating current source. Disturbances of the magnetic field, when a metal particle passes, produces an electrical signal that is measured by the sense coil. The amplitude of the sensor output signal is proportional to the particle mass (Metalscan User's Manual C000833). Sensor output connects to the facility data acquisition computer via an RS232 cable. The sensor measures the number of particles, their approximate size (125 to 1000 μm) and calculates an accumulated mass (Howe and Muir (1998)). Figure 2.4 shows the cross section view of the oil debris sensor. Two filters are located downstream of the oil debris sensor to capture the debris after it is measured by the sensor.

Shaft speed was measured by an optical sensor that creates a pulse signal for each revolution of the shaft. A speed pulse is required for calculation of the vibration algorithms and will be discussed in the next chapter, Vibration Features FM4 and NA4.

Load pressure was measured using a capacitance pressure transducer. Torque on the gear tooth is calculated from this load pressure.

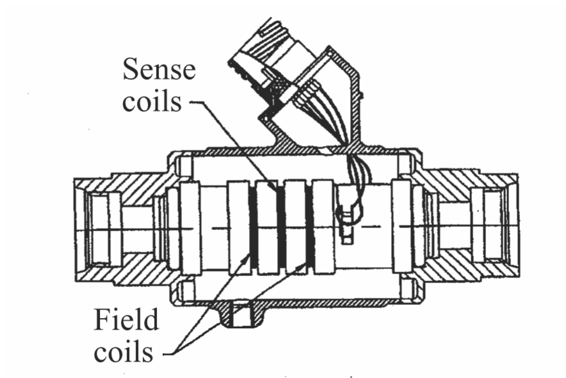


Figure 2.4.—Oil debris sensor cross section.
(Metalscan User's Manual)

Oil debris monitor, speed, pressure, and raw vibration data were collected and processed in real-time using the program ALBERT, Ames-Lewis Basic Experimentation in Real Time, co-developed by NASA Glenn and NASA Ames. ALBERT is a data acquisition program developed to collect data in NASA gear diagnostic test facilities. The program uses a commercially available programming language (Labview™ Basic I Course Manual). ALBERT collects and displays the vibration data from the DAQ card and the oil debris data from the RS232 connection in real-time during tests in the Spur Gear Fatigue Test Rig.

Oil debris and pressure data were recorded once per minute. Reading number, based on data collection rate, is equivalent to minutes and can also be interpreted as mesh cycles equal to reading number times 10^4. Vibration and shaft rotational speed data were sampled at 200 KHz for one second duration every minute. Vibration algorithms FM4 and NA4 were calculated from this data and recorded every minute. The steps to calculate vibration algorithms FM4 and NA4 are discussed in the next chapter, section 3.1.1, Vibration Features FM4 and NA4.

Chapter 3

RESEARCH METHODOLOGY

3.1 Diagnostic Feature Selection and Validation

3.1.1 Vibration Features FM4 and NA4

Two vibration diagnostic parameters were selected as the vibration features for this analysis, FM4 and NA4. FM4 was developed to detect changes in the vibration pattern resulting from fatigue damage on a limited number of teeth (Stewart (1977)). NA4 was developed to detect the onset of fatigue damage and to continue to react to the damage as it spreads (Zakrajsek, et al. (1993)). FM4 and NA4 are dimensionless parameters with nominal values of approximately 3. When gear damage occurs, the values of FM4 and NA4 increase.

Prior to calculating FM4 and NA4, the time-synchronous average of the vibration data is calculated. Synchronous averaging of time signals is a technique used to extract periodic waveforms from additive noise by averaging the vibration signal over one revolution of the shaft. The signal time-synchronous average is obtained by taking the average of the signal in the time domain with each record starting at the same point in the cycle as determined by the once per revolution tachometer signal. Using the above averaging scheme, the desired signal that is synchronous with the shaft rotational frequency will intensify relative to the nonperiodic signals. This time synchronous average signal is used as a basis for FM4 and NA4 methods (Zakrajsek (1989) and Zakrajsek, et al. (1993)).

An example of obtaining the time synchronous average for the vibration data collected for this experiment is shown in Figure 3.1. The plots are displays available from the ALBERT data acquisition software. The first plot shows the raw vibration data sampled at 200KHz for 1 sec duration. The second plot shows one revolution of the 167 cycles averaged. The last plot is the average of 167 revolutions of vibration data interpolated to 1024 points for 1 shaft revolution. This interpolated data is used to calculate FM4 and NA4.

Several statistical and filtering operations are used to calculate FM4. First the regular meshing components are filtered from the signal resulting in a difference signal. The regular meshing components are the shaft and meshing frequencies, their harmonics and first order sidebands. Two statistical operations are then performed on the filtered signal to obtain standard deviation and kurtosis. Kurtosis is the statistical parameter that quantifies how “Gaussian” a time history is, and is defined as the fourth moment of a probability density function (Tustin Technical Institute (1996)).

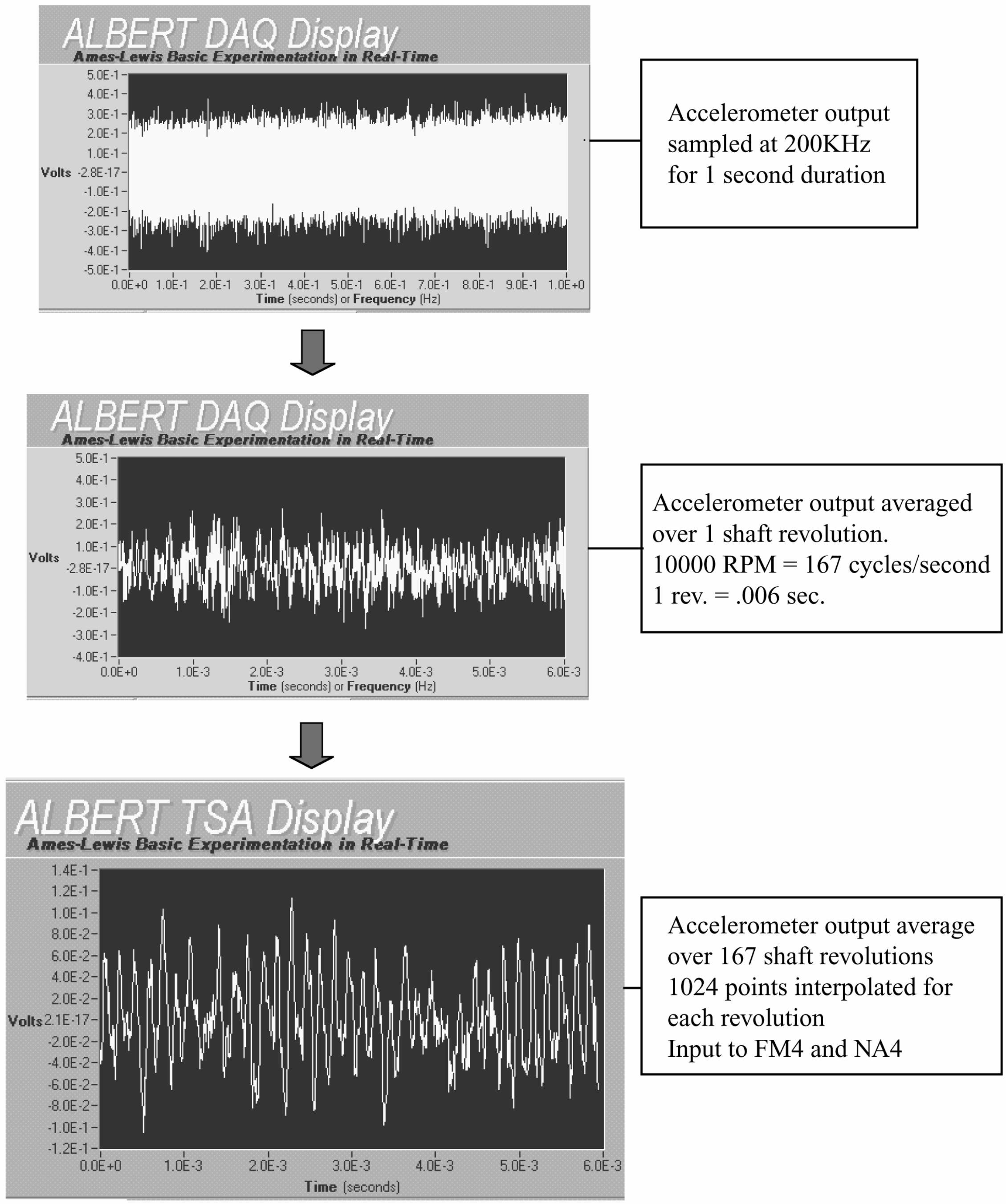


Figure 3.1.—Time synchronous averaging of vibration data.

FM4 is calculated as follows:

$$FM4 = \frac{K}{(RMSDS)^4} \tag{3.1}$$

where K is Kurtosis and RMSDS is the root-mean-square of the difference signal. The Kurtosis is calculated by

$$K = \left[\frac{1}{N} \sum_{i=1}^{N} \left(d_i - \overline{d} \right)^4 \right] \tag{3.2}$$

where d is the difference signal, $\overline{d}$ is the mean value of the difference signal, and N is the total number of interpolated data points per reading. RMSDS, the standard deviation of the difference signal, is calculated by (Zakrajsek (1989))

$$RMSDS = \left[\frac{1}{N} \sum_{i=1}^{N} \left(d_i - \overline{d} \right)^2 \right]^{\frac{1}{2}} \tag{3.3}$$

A flowchart of the calculation procedure is shown in Figure 3.2. Referring to this figure, perform a Fast Fourier Transform on the time synchronous averaged accelerometer data. In the frequency domain, filter the shaft and meshing frequencies of the test gears and slave gears, their harmonics and first order sidebands. Perform an inverse Fast Fourier Transform to return to the time domain. Subtract the filtered signal from the averaged accelerometer data to obtain a difference signal. Calculate the normalized kurtosis and standard deviation of the difference signal. Then, divide kurtosis by standard deviation to the fourth power.

For FM4, the standard deviation of the difference signal indicates the amount of energy in the non meshing components. The kurtosis indicates the presence of peaks in the difference signal. The theory behind FM4 is that for a gear in good condition, the difference signal would be noise with a Gaussian amplitude distribution. The standard deviation should be relatively constant, and normalized kurtosis indicates a value of three. When a tooth develops a major defect, a peak or series of peaks appear in the difference signal, causing the kurtosis value to increase. The standard deviation increases only when the peaks become severe enough to bring up the RMS of the entire difference signal (Zakrajsek (1989)).

The NA4 parameter is calculated in a similar manner to FM4, with two alterations. The first change involves retaining the first order sidebands when filtering the meshing components of the difference signal. The developer of NA4 determined diagnostic information in the sidebands contained useful information and should be maintained (Zakrajsek et al. (1993)). Three plots shown in Figure 3.3 show the NA4 signal in the time and frequency domain before and after filtering. Plot (a) shows the time synchronous averaged signal of the vibration data in the time domain before NA4

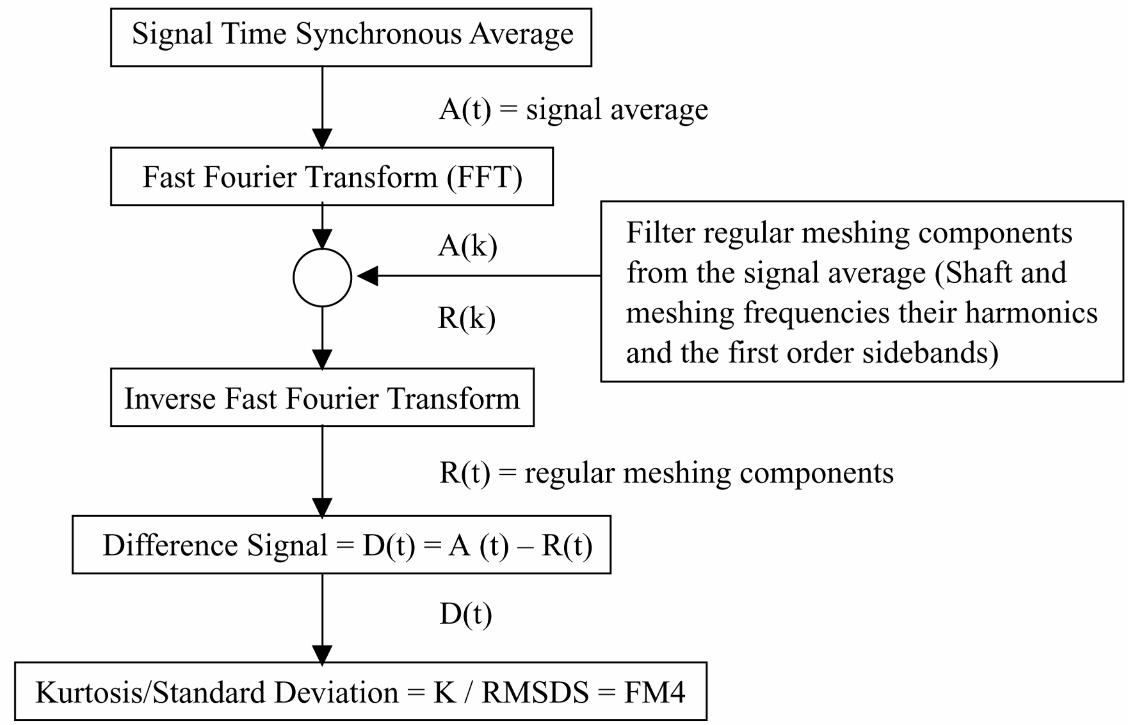


Figure 3.2.—FM4 calculation flowchart. (Stewart (1977)).

filtering. Plot (b) shown the original signal in gray and the filtered part of the signal overlaid in black in the frequency domain. The filtered signals are the test gear and slave gear meshing frequencies (4667 and 5833 Hz) and their harmonics. Plot (c) shows the signal after filtering in the time domain.

The second change is that while FM4 is calculated by the kurtosis of a data record divided by the square of the variance of the same record, NA4 is divided by the square of the average variance. The average variance is the mean value of the variance of all previous data records in the run ensemble (Zakrajsek, et al. (1994b)). The average variance was used to compare changes in the residual signal to the running average of the variance of the system. This caused NA4 to grow with the severity of the fault until the average of the variance itself changes. NA4 is calculated as follows:

$$NA4(M) = \frac{N\sum_{i=1}^{N}(r_i - \bar{r})^4}{\left\{\frac{1}{M}\sum_{j=1}^{M}\left[\sum_{i=1}^{N}(r_{ij} - \bar{r}_j)^2\right]\right\}^2} \tag{3.4}$$

where r is the residual signal obtained by removing shaft frequencies, meshing frequencies, and their harmonics from the FFT of the time synchronous averaged signal, $\bar{r}$ is the mean value of residual signal, N is the total number of interpolated data points per reading, *i* is the interpolated data point index per reading, M is the current reading number (total number of data points for averaging), and *j* is the reading number.

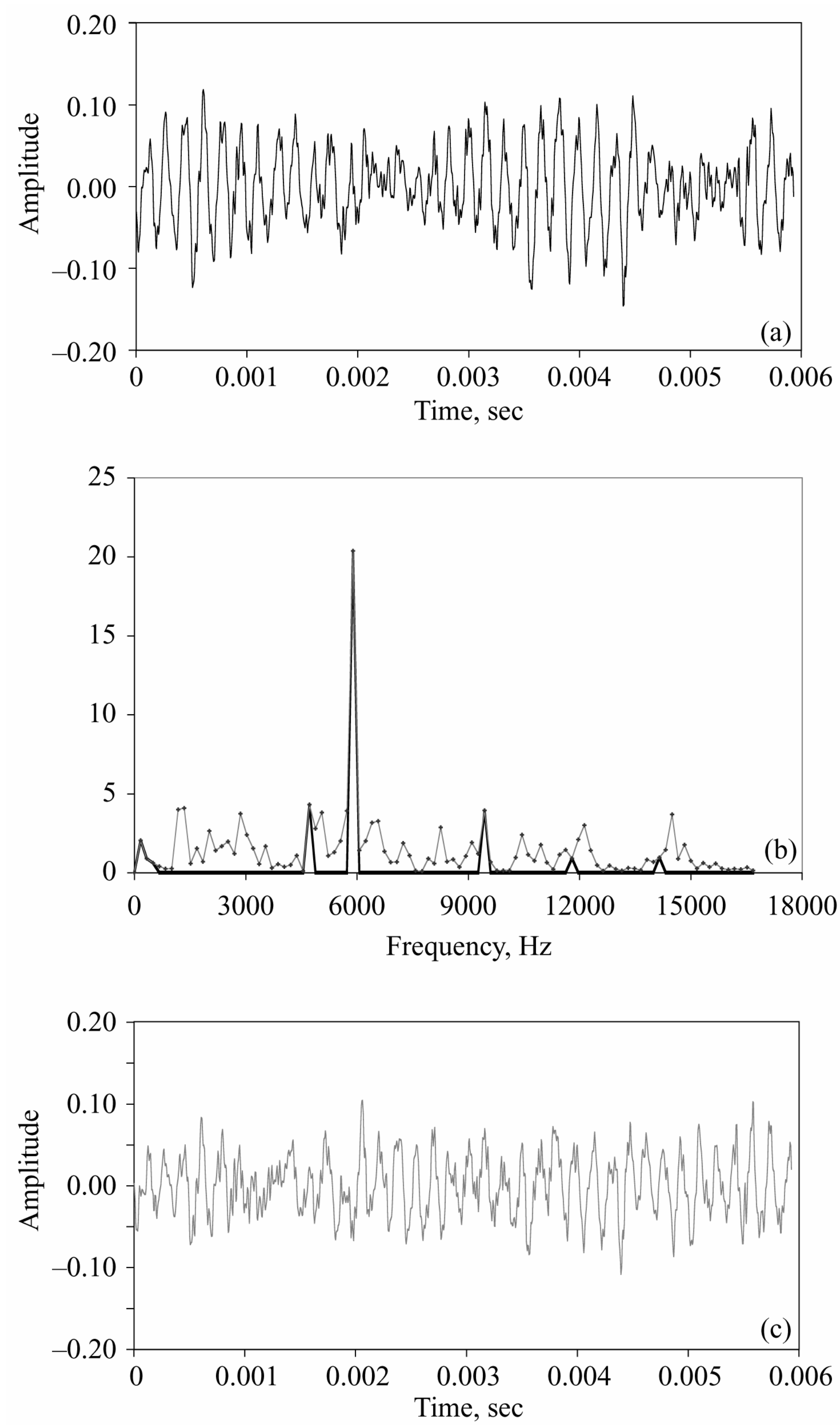


Figure 3.3— (a) Time synchronous averaged signal. (b) Converted to the frequency domain with meshing components filtered. (c) After NA4 filtering.

3.1.2 Preliminary Evaluation of Damage Detection Features

Preliminary tests were performed in the Spur Gear Fatigue Test Rig. The objective of these tests was to assess the capability of the individual parameters to detect gear pitting damage. If the parameters were unable to predict transmission health separately, very little benefit would be obtained by fusing them together. The original focus of these preliminary tests was to determine if the oil debris sensor was sensitive enough to detect pitting fatigue failures of gears. Although the two vibration algorithms were selected based on their past history of successfully predicting gear pitting damage, it was found improvements were required to the vibration algorithms prior to integration. Results of this preliminary study are discussed in this section.

The analysis is based on data collected from two gear tests that ended when pitting damage occurred. Figure 3.4 is a plot of the data measured during testing of Gear Set 1. Vibration algorithms FM4, NA4, and the accumulated mass measured by the oil debris monitor (ODM), referred to as the oil debris sensor, are plotted versus reading number. Readings were recorded once per minute. This test collected 13716 readings over 228 hours. FM4 and NA4 were calculated for both the accelerometer located on the shaft and the accelerometer located on the housing. During the 228 hours of testing, ten shutdowns occurred. To restart after shutdown, the rig was brought up to speed, and load was reapplied. These load changes caused significant spikes in the NA4 plot that can be readily observed on Figure 3.4 following shutdowns at readings 1455, 2576, 3663, 3736, 3982, 4128, 4681, 5035, 5309, and 5435. The sensitivity to load was due to the changes of the running average in the denominator of this algorithm. Unfortunately, this change was due to a load change, not a damaged gear.

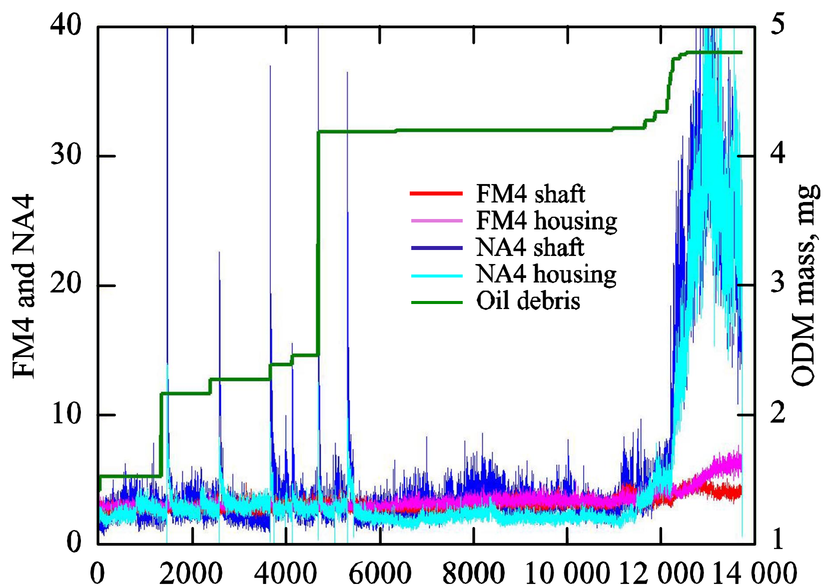


Figure 3.4.—Vibration and oil debris data for gear set 1.

The sensitivity of NA4 to even minor changes in load has been documented in several research papers (Zakrajsek, et al. (1994a) and Zakrajsek, et al. (1995a)). NA4 Reset was developed as the result of these initial findings and is discussed in detail in section 3.1.3, Justification of Vibration Feature NA4 Reset. Another observation to note on Figure 3.4 is that after the shutdown at reading 4681, the oil debris monitor indicated one 725 to 775 micron particle passed through the sensor, causing a large increase in the accumulated mass. This one large chip was apparently flushed out of the line when the rig was restarted after the shutdown. This finding established the large role operational effects play on the diagnostic tools and the need to take these effects into consideration when developing a reliable health monitoring system.

Initial pitting appeared to occur at reading 11647. At the completion of the test, the gears were inspected for damage. Initial pitting was observed on tooth 12 of both the driven and driver gears. By visual observation of the overall plot on Figure 3.4, all parameters showed a significant increase when pitting damage began to occur. Figure 3.5 has an expanded scale in order to observe the increase in NA4 and FM4 as pitting damage progressed. Reviewing Figures 3.4 and 3.5, vibration algorithms FM4 and NA4 for both accelerometers, and the accumulated mass increase significantly when pitting damage occurs. Figure 3.6 shows photos of the damage on the driver and driven tooth 12 at the completion of the test. Damage progression images were not obtained for these experiments because the video inspection system was installed after this preliminary evaluation.

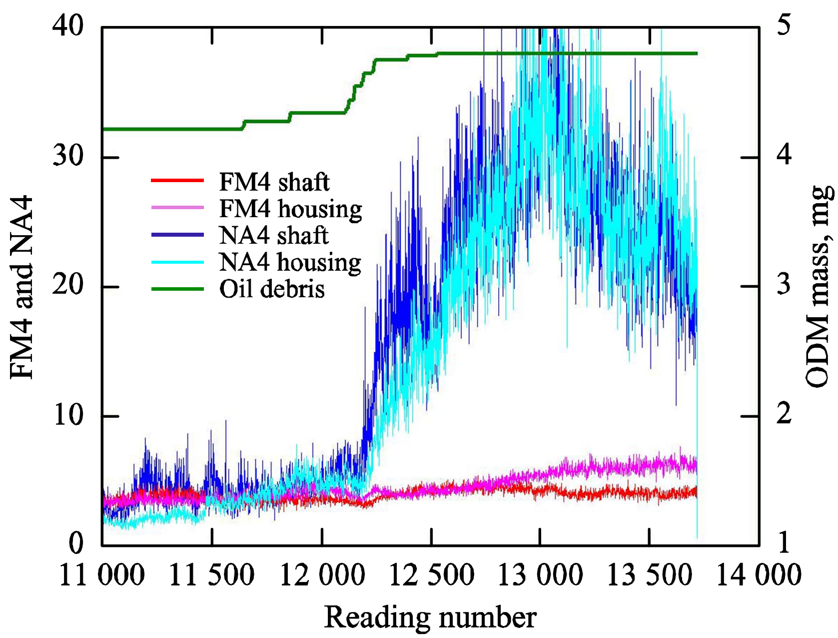


Figure 3.5.—Vibration and oil debris data for gear set 1 (expanded scale).

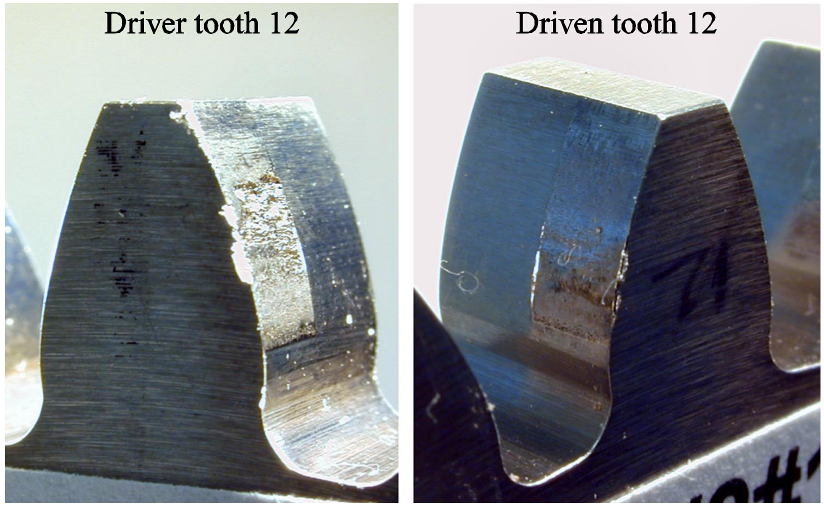


Figure 3.6.—Gear damage at completion of gear set 1 test.

Figures 3.7 and 3.8 are plots of the data measured during testing of Gear Set 2. Vibration algorithms FM4, NA4, and the accumulated mass measured by the oil debris monitor are plotted versus reading number. Readings were recorded once per minute. During this test, 5314 readings were collected over 88 hours. FM4 and NA4 were calculated for both the accelerometer located on the shaft and the accelerometer located on the housing. Initial pitting appeared to occur at reading 5020. Gears were inspected at Reading 5181 and initial pitting was observed on teeth 15 and 16 of the driver gear and teeth 15,16, and 17 of the driven gear. The gears were inspected at the completion of the test. From the inspection, initial pitting was observed on driver teeth 19,24, and 27, and driven tooth 14. A combination of initial and destructive pitting was observed on both the driver and driven gears teeth numbers 9,15,16,17,18, and 24. Figure 3.9 shows photos of the damaged teeth at test completion.

Referring to Figures 3.7 and 3.8, all parameters show a significant increase when pitting damage occurs. A shutdown with a load change at reading 380 caused the large spike of NA4. Shutdowns with load fluctuations also occurred at readings 4903, 4919, 5128, and 5180. As shown on Figure 3.8, after the shutdown at Reading 4919, FM4 and NA4 increased then decreased slightly. This increase/decrease is most likely due to the load change. From these initial tests, an oil debris feature using the step change of mass over the time from the last step change of mass was looked at as a potential indicator of gear damage. However, the tests performed in the Spur Rigs are accelerated tests. Deriving a relationship between this time dependent oil debris feature and tests performed on other systems would be impossible. Further testing indicated accumulated oil debris provided better results for detecting gear pitting damage. The oil debris mass feature selected will be discussed in section 3.1.4, Oil Debris Feature.

From this preliminary evaluation it was determined that defining specific threshold limits for vibration algorithms to indicate when pitting damage has occurred is a challenging task. Values of FM4 and NA4 are non-dimensional numbers with nominal values of 3. The threshold values vary when damage occurs. Several research papers define a range from 7 to 15 as NA4 reacts to pitting damage (Zakrajsek, et al. (1993) and

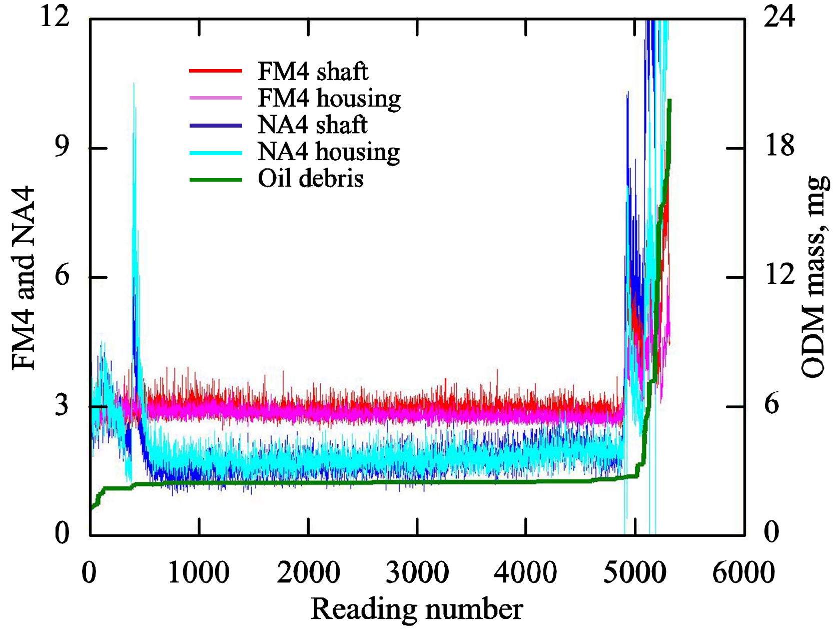


Figure 3.7.—Vibration and oil debris data for gear set 2.

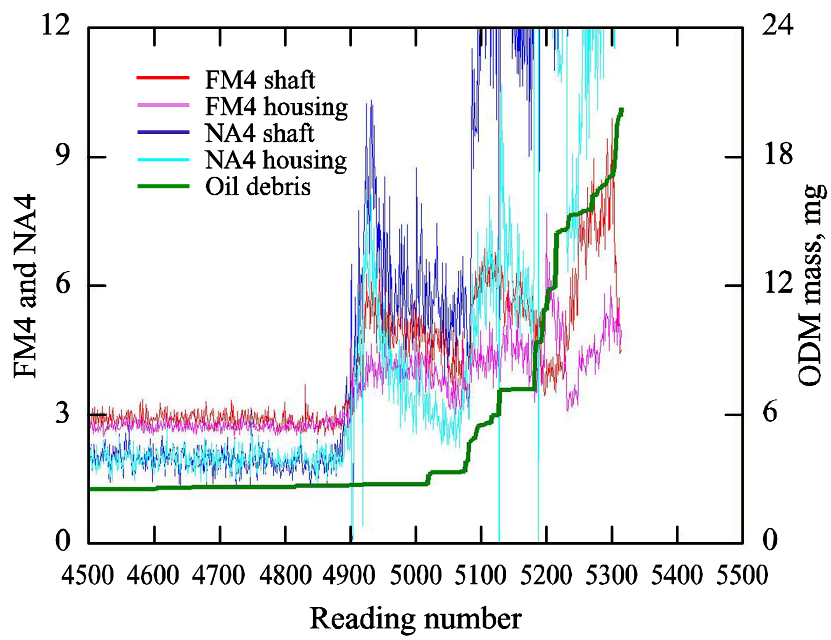
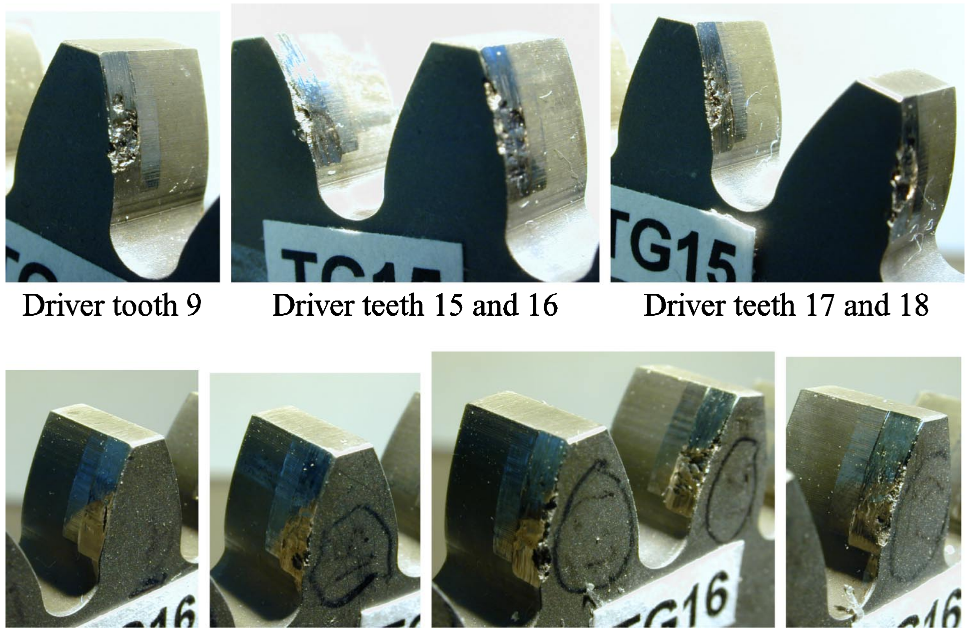


Figure 3.8.—Vibration and oil debris data for gear set 2 (expanded scale).



Driver tooth 9 Driver teeth 15 and 16 Driver teeth 17 and 18

Driven tooth 9 Driven tooth 15 Driven teeth 16 and 17 Driven tooth 18

Figure 3.9.—Gear damage at completion of gear set 2 test.

Choy et al. (1994)). For parameter FM4, values for initiation of pitting ranged from 3 to 7 (Zakrajsek, et al. (1993, 1995)). One simple method for setting threshold limits is to set the limit by adding 3 times the standard deviation to the mean of baseline vibration data (Slemp and Skeirik (1999)). During this preliminary evaluation, three additional tests were run on the test rig that generated no damage on the test gears. The run hours ranged from 350 to 497 hours for each test with a total of 1204 hours. The data recorded for FM4 and NA4 during the tests when no damage occurred was used to apply this simple method to set threshold limits for these algorithms. This was done by calculating the mean and standard deviation during each test, and adding 3 times the standard deviation to the mean. Because the number of readings for each test varied, a weighted average of the limit was calculated based on the number of readings recorded during each test. From this exercise the limit for FM4 was 4.45 and the limit for NA4 was 7.33. Based on these threshold limits, FM4 indicated pitting damage sooner than NA4. NA4 had the most false alarms for this preliminary evaluation due to the sensitivity of NA4 to the load changes. A different technique for setting threshold limits to minimize false alarms will be discussed in subchapter 3.3, Feature Validation for Sensor Fusion.

This preliminary research assessed the reliability of the individual parameters when detecting gear pitting. Based on the data collected, FM4, NA4, and the oil debris mass each showed a significant increase when pitting damage began to occur. Improving the reliability of the individual parameters must be done prior to integrating the three parameters into an intelligent health monitoring methodology. Results of this research identified several improvements to be made to the parameters to increase their individual performance. The first improvement, decrease the sensitivity of NA4 to load changes, resulted in the development of NA4 Reset, to be discussed in section 3.1.3, Justification

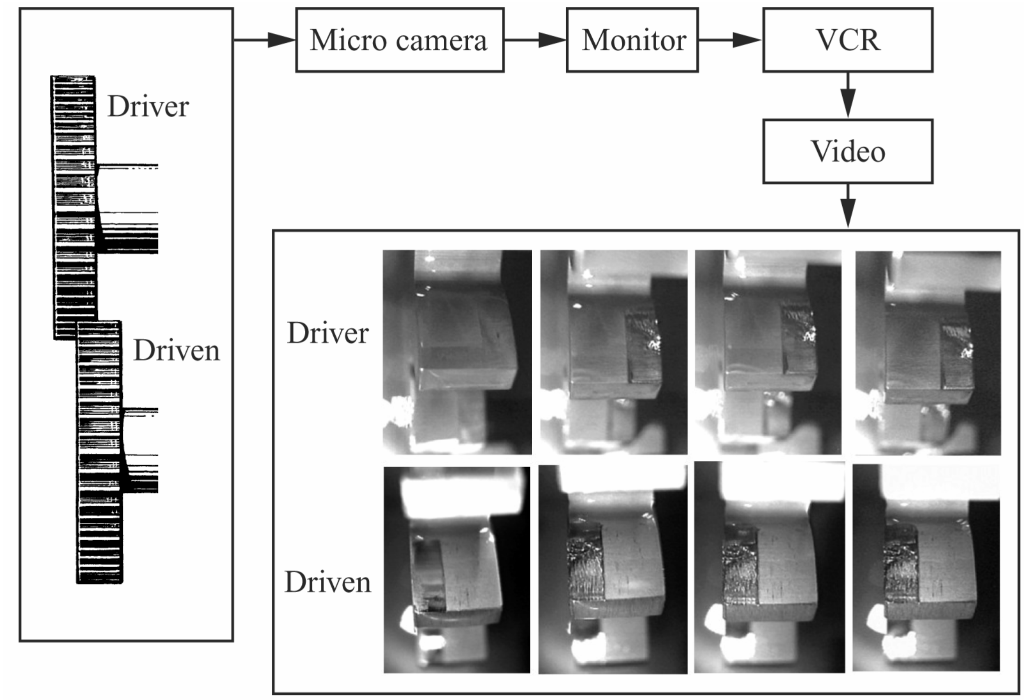


Figure 3.10.—Video inspection system.

of Vibration Feature NA4 Reset. The second improvement was to develop a system for documenting damage progression. This resulted in the installation of the video inspection system for damage progression documentation. Figure 3.10 is a block diagram showing the video inspection system. A micro camera is inserted in the gearbox viewing ports to view the driver and driven gear tooth contact area and to record these images on a VCR. The third improvement was to define a method to set alert and fault threshold limits for all three parameters. This method will be discussed in the section 3.3, Feature Validation for Sensor Fusion. It should be noted that Gear Set 1 and 2 herein will be referred to as Experiments 7 and 8.

3.1.3 Justification of Vibration Feature NA4 Reset

As discussed in Chapter 3, section 3.1.1, Vibration Features FM4 and NA4, when gear pitting starts, the magnitude of NA4 shows a significant increase above its typical value of 3. Unfortunately, as observed in the data from the preliminary evaluation, NA4 responds similarly to load changes. The sensitivity of NA4 to even minor changes in load has been documented in several research papers (Zakrajsek, et al. (1994a) and Zakrajsek, et al. (1995a)). The magnitude of NA4 reacts to changes in load since the load change affects the running average in the denominator of this algorithm. When using this algorithm to detect gear pitting damage on helicopter gearboxes in different flight regimes, the load effect on this algorithm must be minimized. Preliminary experiments have shown the need to minimize the effect of load on NA4. Vibration feature NA4 Reset was developed to minimize these load effects and will be discussed in this chapter.

A change to the calculation of NA4 was required to minimize the effect of a fluctuating load on NA4. This change, NA4 reset, is made when the load increases or decreases by a given percentage of the nominal value. For this application, a 10 percent load change limit was used. This 10 percent limit was used after analyzing the data to determine the largest load fluctuation NA4 could withstand without increasing in magnitude. If the system to apply NA4 had continuous load fluctuations, the benefit of maintaining a running average in the denominator would be minimal. For NA4 reset, when the load changes by 10 percent, the denominator resets to the square of the variance of the same reading, and a new average variance for this load is calculated starting with the reading measured when the load changed. Each time the load changes beyond 10 percent, the first reading in the average variance resets to the first reading when the load changed. This first reading is calculated as follows:

$$NA4 = \frac{N\sum_{i=1}^{N}(r_i - \bar{r})^4}{\left[\sum_{i=1}^{N}\left(r_i - \bar{r}\right)^2\right]^2} \tag{3.5}$$

The denominators for the readings that follow this load change are calculated as the square of the average variance, the mean value of the variance of all previous readings starting with the first reading when the load changed. Each time the load changes by 10 percent, the denominator is reset by using equation (3.5) for the initial reading.

The analysis discussed in this section is based on data collected during 4 experiments in the NASA Glenn Spur Gear Fatigue Rig, 3 of which involved pitting damage. All of the instrumentation discussed in Chapter 2, Experimental Setup and Procedures, was collected during these experiments. NA4 was calculated for both accelerometers, and responded similarly, but only the accelerometer on the shaft is plotted for this analysis. The first experiment was to verify the effect of load on the NA4 parameter. The load was increased and decreased with NA4 calculated from the vibration data. The gear set had no evidence of pitting before or after the test. A plot of Load Pressure, NA4 and NA4 Reset, for the first experiment, is shown on Figure 3.11. Data were collected every minute, so reading number is equivalent to minutes. Since the shaft speed is 10,000 rev/min, reading number can also be interpreted as mesh cycles equal to reading number times 10^4.

As discussed previously, NA4 reset is the same as NA4 except the average variance in the denominator is reset each time the load fluctuated by 10 percent. From this plot, the sensitivity of NA4 to changes in load can be easily observed. NA4 appeared to track load pressure. The plot of NA4 reset in Figure 3.11 shows that applying this technique minimizes the sensitivity of NA4 to load.

Although the sensitivity of NA4 to load changes can be corrected by resetting the denominator, one must verify that applying this technique does not significantly decrease the sensitivity of NA4 to pitting damage. Data from 3 experiments when pitting damage occurred and the load fluctuated were used to verify resetting the denominator of NA4 did not decrease its sensitivity to pitting damage. Descriptions of the pitting damage that occurred during these 3 experiments are listed in Tables 3.1 to 3.3. Photos of damage

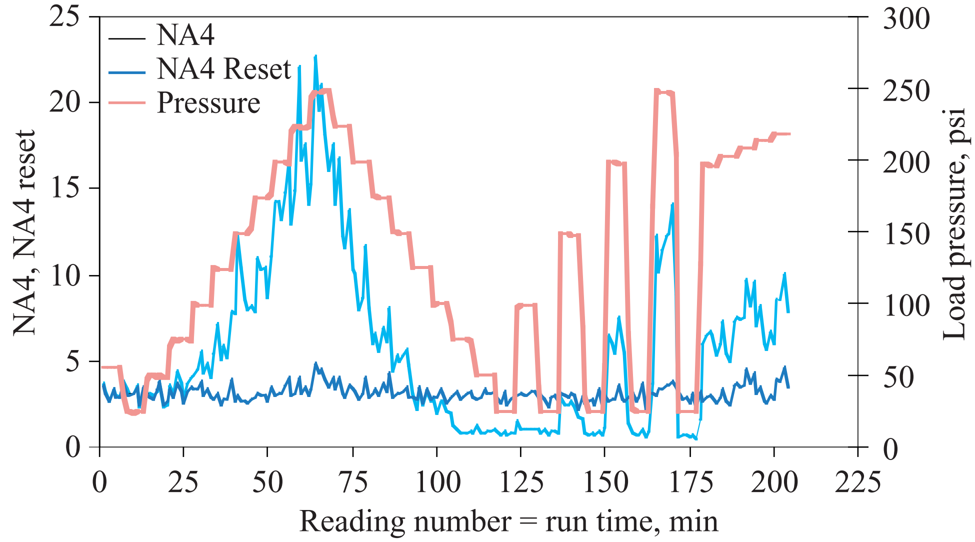


Figure 3.11.—Data from experiment illustrating load effects.

progression on a selected tooth during each experiment are shown on Figures 3.12 to 3.14. The test gears are run offset to provide a narrow effective face width to maximize gear contact stress and minimize bending fatigue failures. Damage levels for this analysis are defined as follows:

1. **Wear**: Layers of metal uniformly removed from the surface
2. **Initial Pitting**: Pits less than 1/64 in. in diameter and cover less than 25 percent of tooth contact area
3. **Destructive Pitting**: Pits greater than 1/64 in. in diameter and cover greater than 25 percent of tooth contact area

Initial pitting on specific teeth will only be discussed in reference to test completion. Although initial pitting most likely occurred prior to test completion, a detailed analysis of the inspection images is required to verify when it occurred and is outside the scope of this research.

Plots of the data measured during these 3 experiments are shown on Figures 3.15 to 3.21. Two different plots are shown for each experiment. The first plot is of Load Pressure, NA4 and NA4 reset for each experiment. The diamonds indicate when the rig was restarted after a shutdown. The second is a plot of FM4, NA4 Reset and the accumulated mass from the oil debris monitor. The triangles on the X-axis indicate the reading number that the rig was shutdown for inspection. These reading numbers are listed in tables 3.1 to 3.3. Each experiment will be discussed in turn.

TABLE 3.1
Damage Description for Experiment 1

Reading Number Run Time (min)	Damage Description	Teeth damaged on Driver Gear	Teeth damaged on Driven Gear
60	Run-in Wear	All	All
120	Wear	All	All
1581	Wear	All	All
10622	Wear	All	All
14369	Wear Destructive Pitting	All 6	All 6
14430	Wear Destructive Pitting	All 6	All 6
14512	Wear Destructive Pitting	All 6,7	All 6,7
14688	Wear Destructive Pitting	All 6,7	All 6,7
14846	Wear Destructive Pitting	All 6,7	All 6,7
15136	Wear Initial Pitting Destructive Pitting	All teeth All teeth 6,7,8	All teeth 6,7,10

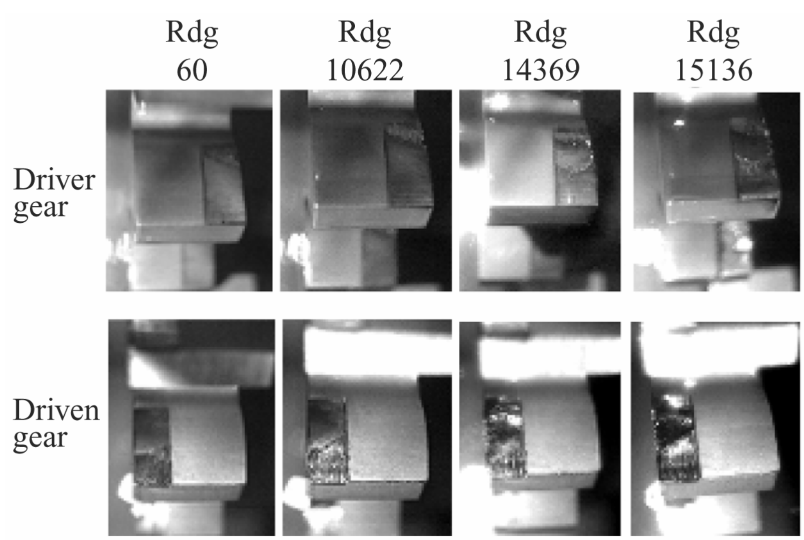


Figure 3.12.—Damage progression of driver/driven tooth 6 for experiment 1.

TABLE 3.2
Damage Description for Experiment 2

Reading Number Run Time (min)	Damage Description	Teeth Damaged on Driver Gear	Teeth Damaged on Driven Gear
1573	Run-in Wear	All	All
2199	Wear Destructive Pitting	All	All 11
2296	Wear Destructive Pitting	All	All 10, 11
2444	Wear Initial Pitting Destructive Pitting	All All 10,11	All 10, 11, 14 10,11,14

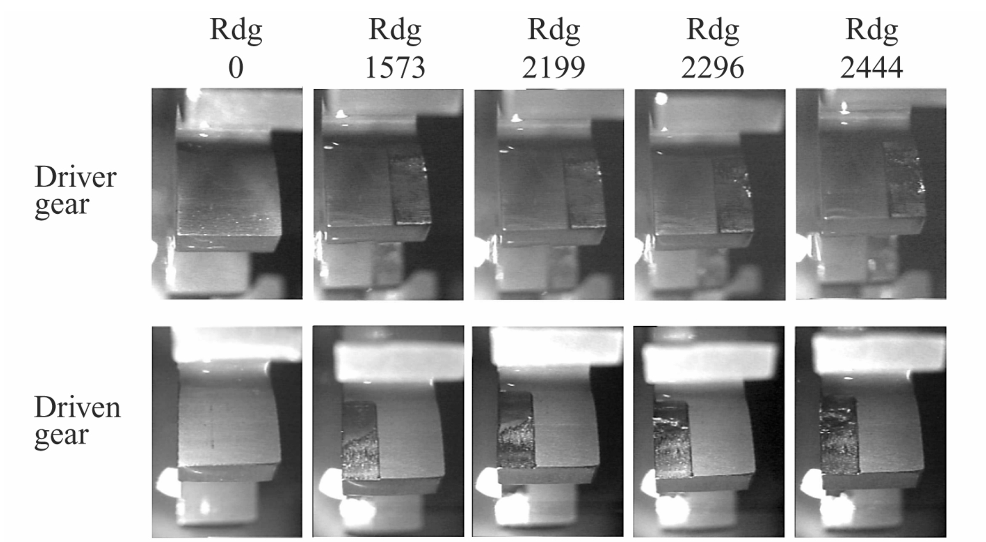


Figure 3.13.—Damage progression of driver/driven tooth 11 for experiment 2.

TABLE 3.3
Damage Description for Experiment 3

Reading Number Run Time (min)	Damage Description	Teeth Damaged on Driver Gear	Teeth Damaged on Driven Gear
58	Run-in Wear	All	All
2669	Wear Destructive Pitting	All 1,28	All 1,28
2857	Wear Destructive Pitting	All 1,6,28	All 1,6,28
3029	Wear Initial Pitting Destructive Pitting	All All 1,6,28	All 1,6,28 1,6,28

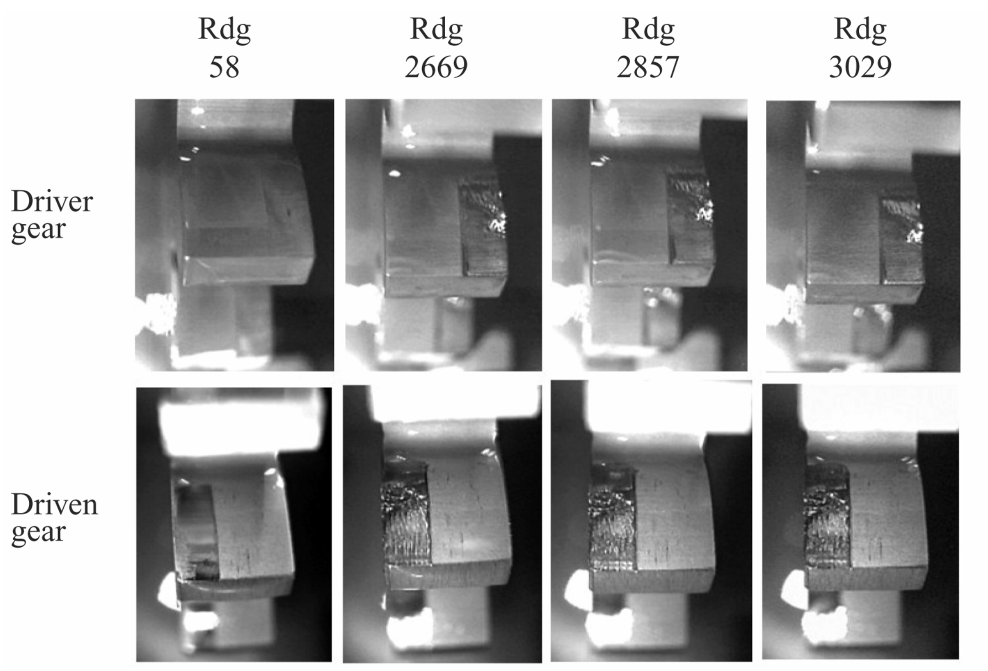


Figure 3.14.—Damage progression of driver/driven tooth 28 for experiment 3.

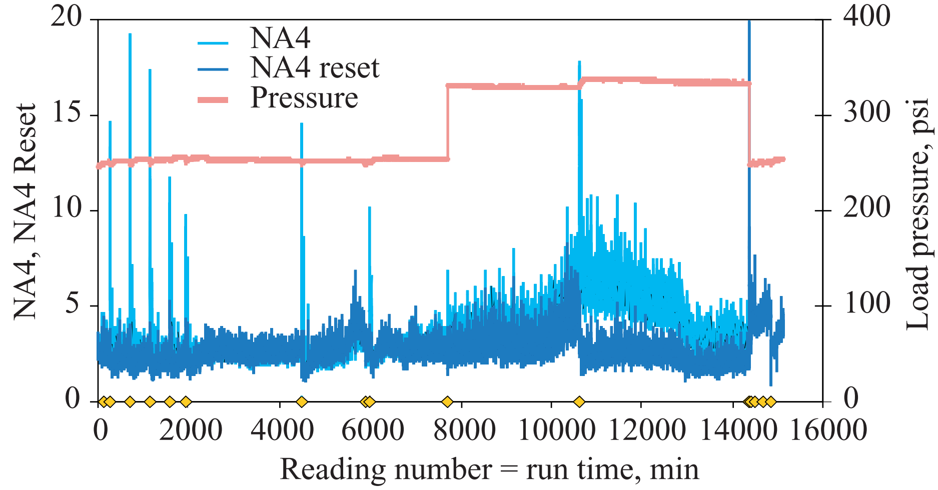


Figure 3.15.—Data from experiment 1 illustrating load effects.

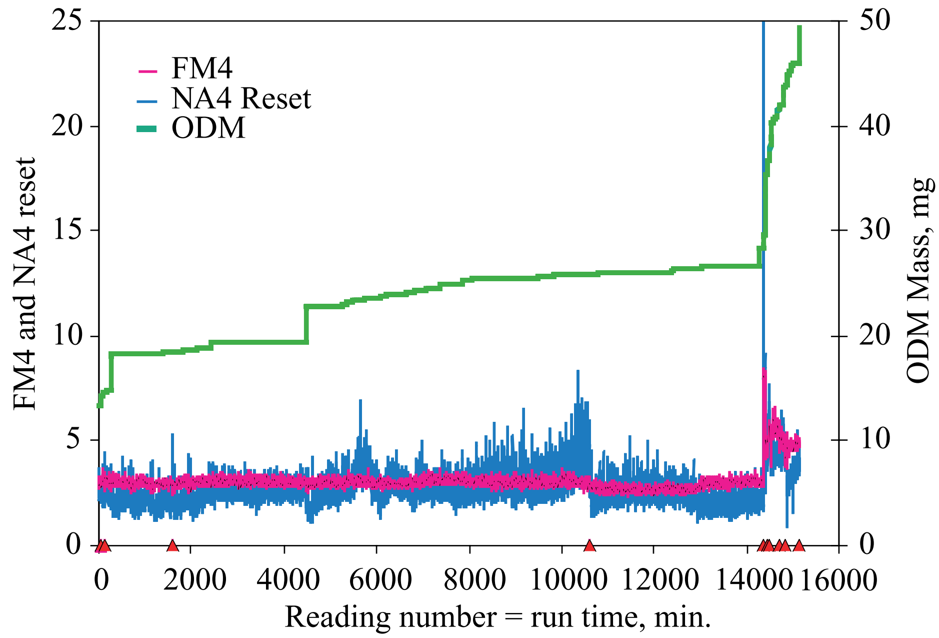


Figure 3.16.—Vibration, ODM, and damage data from experiment 1.

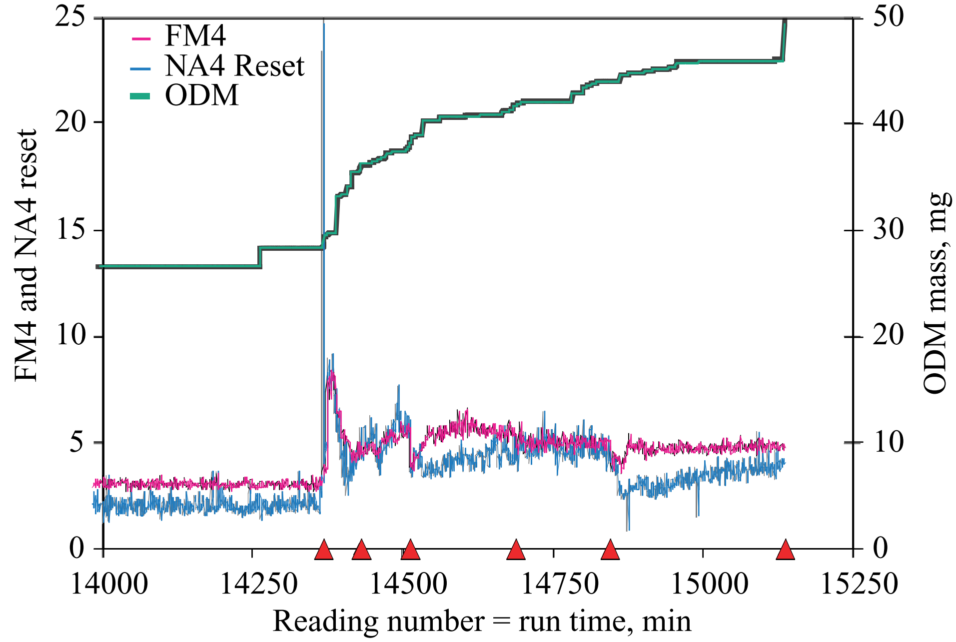


Figure 3.17.—Vibration, ODM, and damage data from experiment 1.

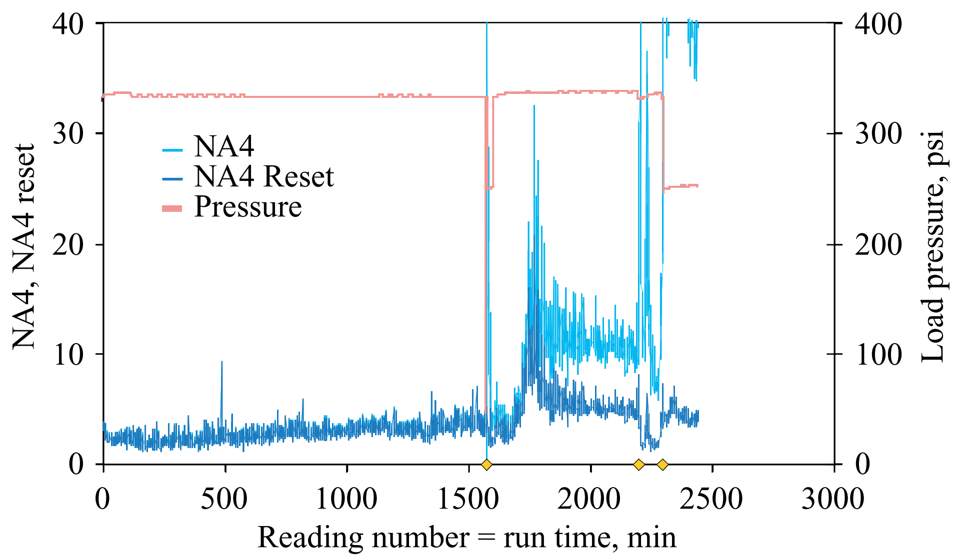


Figure 3.18.—Data from experiment 2 illustrating load effects.

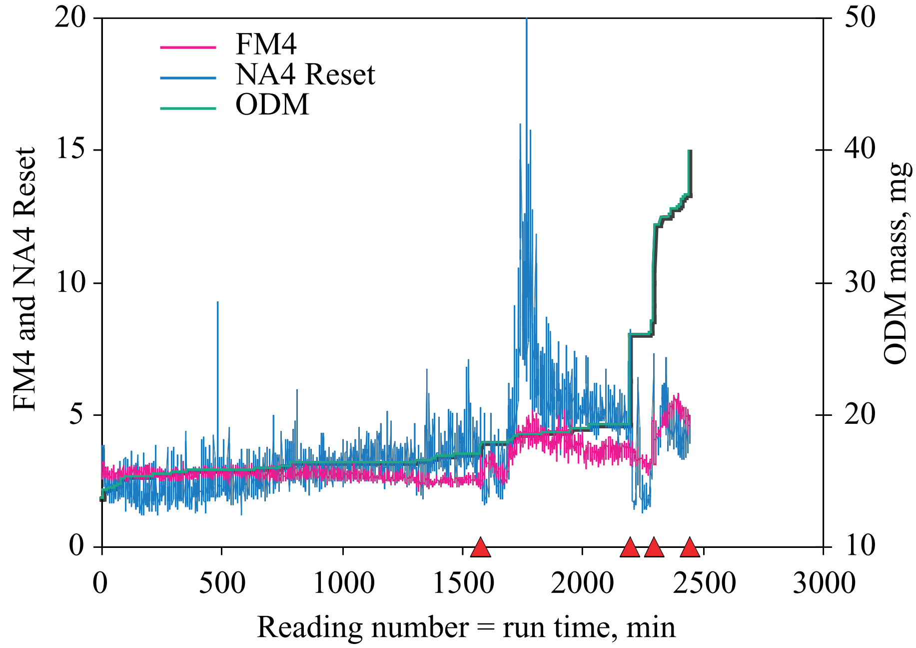


Figure 3.19.—Vibration, ODM, and damage data from experiment 2.

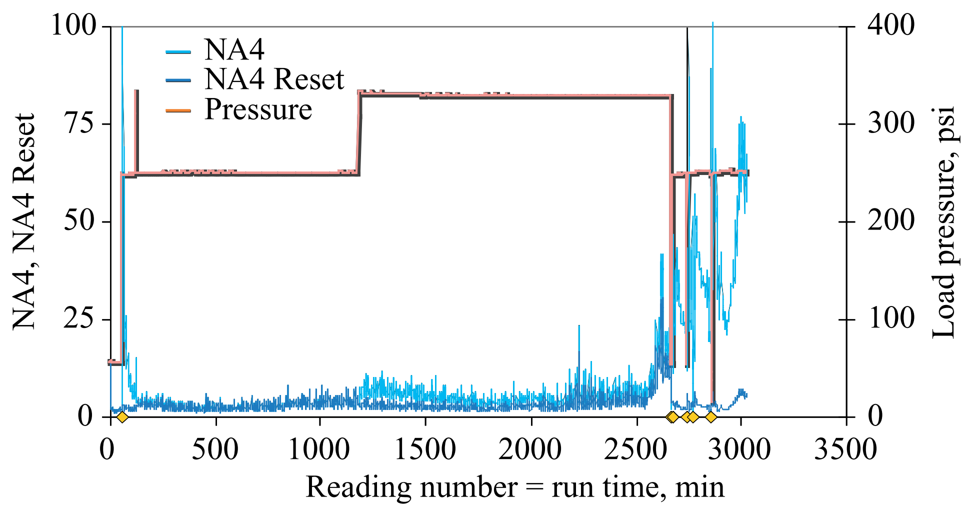


Figure 3.20.—Data from experiment 3 illustrating load effects.

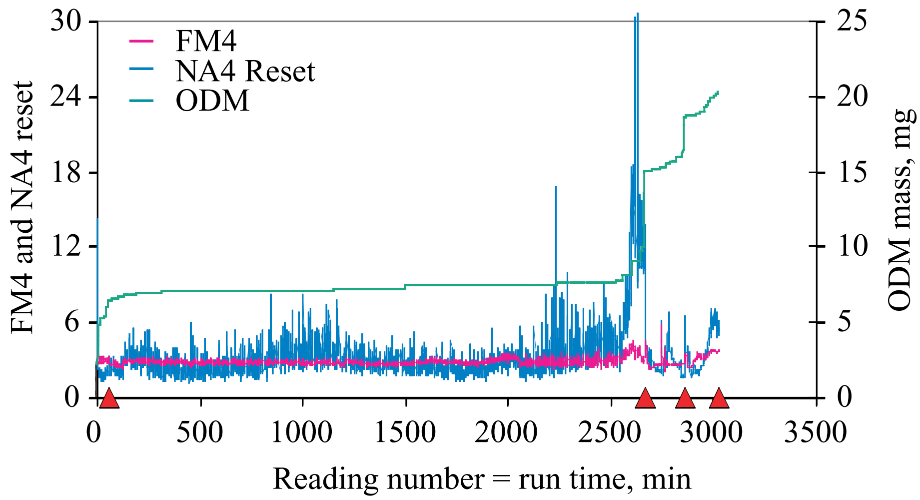


Figure 3.21.—Vibration, ODM, and damage data from experiment 3.

Results from Experiment 1 are plotted in Figures 3.15 to 3.17. Figure 3.15 shows the effect of the rig restarts after shutdowns on NA4 by the NA4 magnitude spikes that occur after shutdowns. Figures 3.16 and 3.17 indicate damage occurred just prior to inspection at reading 14369. Inspection at reading 14369 indicated destructive pitting first occurred on driver and driven tooth 6. The progression of damage is detailed in Table 3.1 and Figure 3.12. NA4 and FM4 both indicate an increase in magnitude when it appears destructive pitting occurred. NA4 Reset, like FM4, is less sensitive to damage as it progresses to a number of teeth and becomes more severe.

Figures 3.18 and 3.19 show plots of data obtained from experiment 2. Damage progression is shown in Table 3.2 and Figure 3.13. Destructive pitting occurred on driven tooth 11 prior to inspection at reading 2199. From Figure 3.18, FM4 and NA4 both indicate and increase in magnitude at approximately reading 1700. As seen previously, both become less sensitive to damage as it progresses.

Results from experiment 3 are plotted in Figures 3.20 and 3.21. Damage progression is shown in Table 3.3 and Figure 3.14. Destructive pitting occurred on driver and driven teeth 1 and 28 prior to inspection at reading 2669. From Figure 3.21, FM4 and NA4 both indicate an increase in magnitude prior to inspection at reading 2669 and become less sensitive to damage as it progresses.

As seen in Figures 3.15 to 3.21, NA4 does react to pitting damage. However, some of the response magnitude is lost with the reset operation. NA4 reset does increase the stability of the NA4 parameter thus enabling it have a more consistent threshold limit. This is a key factor in reducing false alarm rates.

The goal of this part of the research was to define relevant meaningful vibration features prior to integration into a health monitoring system. During the preliminary evaluation, it was found that one of the selected vibration algorithms, NA4, does not perform well under varied load conditions. In order to maintain the integrity of this algorithm, a new algorithm was developed to minimize the effect of load, while maintaining its sensitivity to pitting damage. This new algorithm is referred to as NA4 Reset. Results indicate NA4 Reset is no longer sensitive to load changes, but is still sensitive to pitting damage (Dempsey and Zakrajsek (2001)). Both NA4 Reset and FM4 indicate when destructive pitting occurs on one gear tooth. NA4 Reset, like FM4, is less sensitive to damage as it progresses to a number of teeth and increases in severity. The magnitude of NA4 Reset is less than NA4 when pitting damage occurs, requiring a smaller threshold limit to indicate pitting damage. However, the magnitude of NA4 Reset is significantly larger than FM4 when pitting damage begins to occur. All NA4 data collected for the purpose of this research were recalculated to NA4 Reset prior to integration.

3.1.4 Oil Debris Feature

Part of this research project was to identify the best feature for detecting gear pitting damage from a commercially available on-line oil debris sensor. This was done using oil debris data collected from 9 experiments with no damage and 8 with pitting damage in the NASA Glenn Spur Gear Fatigue Rig. Oil debris feature analysis was performed on this data. Damage progression data (video images) were also collected from 6 of the experiments with pitting damage. During each test, data from an oil debris sensor were monitored and recorded for the occurrence of pitting damage. The data measured from the oil debris sensor during experiments with damage and with no damage were used to identify membership functions to build a simple fuzzy logic model. Using fuzzy logic techniques and the oil debris data, threshold limits were defined that discriminate between stages of pitting wear.

The oil debris sensor records counts of particles in bins set at particle size ranges. The particle size is measured in microns. For these experiments, 16 bins were defined. The range of the bin sizes in microns is shown in Table 3.4. Based on the bin configuration, the average particle size for each bin is used to calculate the cumulative mass for the experiment. This average particle size for each bin is also listed in Table 3.4. The shape of the particle is assumed to be a sphere with a diameter equal to the average particle size. An approximate density of 7922 kg/m^3 is used to calculate the accumulated mass.

TABLE 3.4
Oil debris particle size ranges

Bin	Bin range, µm	Average	Bin	Bin range, µm	Average
1	125–175	150	9	525–575	550
2	175–225	200	10	575–625	600
3	225–275	250	11	625–675	650
4	275–325	300	12	675–725	700
5	325–375	350	13	725–775	750
6	375–425	400	14	775–825	800
7	425–475	450	15	825–900	862.5
8	475–525	500	16	900–1016	958

Experiments 1 to 6 were performed with pitting damage and the video inspection system installed on the rig. Table 3.5 lists the reading numbers when inspection was performed and the measured oil debris mass at this reading. The highlighted cells for each experiment identify the reading number and the mass measured when destructive pitting was first observed on one or more teeth. As can be seen from this table, the amount of mass measured varied significantly for each experiment. The amount of damage to the gear teeth also varied significantly. For example, destructive pitting was observed on 2 teeth for experiment 1, and 4 teeth for experiment 3, yet the debris measured during experiment 3 was significantly lower than the debris measured during experiment 1. This was due to different levels of wear and initial pitting that were distributed on the teeth but were difficult to measure quantitatively using the video inspection system.

Experiments 7 and 8 were performed with visual inspection and both experiments had pitting damage. Table 3.6 lists the reading numbers when inspection was performed and the measured oil debris mass at this reading. Only initial pitting occurred during experiment 7. Initial pitting was observed at reading 5181 for experiment 8, and destructive pitting at reading 5314.

Experiments 9 to 17, wherein no damage was observed at test completion, are listed in Table 3.7. At the completion of experiment 10, 5.453 mg of debris was measured, yet no damage occurred. This is more than the debris measured during experiment 7 (3.381 mg) when initial pitting was observed. This and observations made from the data collected during experiments when damage occurred made it obvious that simple linear correlations could not be used to obtain the features for damage levels from the oil debris data.

TABLE 3.5
Reading Numbers and Oil Debris Mass at Video Gear Inspection.

Experiment 1		Experiment 2		Experiment 3		Experiment 4		Experiment 5		Experiment 6	
Rdg#	Mass (mg)	Rdg#	Mass (mg)	Rdg#	Mass (mg)	Rdg#	Mass (mg)	Rdg#	Mass (mg)	Rdg#	Mass (mg)
60	1.003	1573	3.285	58	0	64	0	62	0	60	0
120	1.418	2199	8.934	2669	8.69	150	2.233	1405	4.214	2810	3.192
1581	5.113	2296	16.267	2857	11.889	378	8.297	2566	7.413	2885	6.396
10622	12.533	2444	26.268	3029	14.148	518	9.462	4425	10.811	2957	8.704
14369	15.475					2065	12.132			9328	11.692
14430	22.468					2366	13.977			12061	14.365
14512	24.586					3671	17.361			12368	22.851
14688	28.451					4655	23.12				
14846	30.686					4863	26.227				
15136	36.108										

*Note: Highlighted cells identify reading and mass when destructive pitting was first observed

TABLE 3.6
Oil Debris during experiments
with visual inspection.

Experiment 7		Experiment 8		Pitting Damage
Rdg#	Mass, mg	Rdg#	Mass, mg	
13716	3.381	5181	6.012	Initial
		5314	19.101	Destructive

TABLE 3.7
Oil Debris at completion of experiments with no damage.

Experiment	Readings	Oil Debris Mass (mg)
9	29866	2.359
10	20452	5.453
11	204	0.418
12	15654	2.276
13	25259	3.159
14	5322	0
15	21016	0.125
16	380	0.099
17	21066	0.064

Prior to discussing methods for feature extraction, it may be beneficial for the reader to get a feel for the amount of debris measured by the oil debris sensor and the amount of damage to one tooth. Applying the definition of destructive pitting, 25 percent of contact surface area for one tooth for these experiments is approximately 0.04322 cm^2. A 1/64 in. (0.0397 cm) diameter pit, assumed spherical in size is equivalent to 0.26 mg oil debris mass. This mass is calculated based on the density used by the sensor software to calculate mass. If 0.0397cm diameter pits densely covered 25 percent of the surface area of 1 tooth, it would be equivalent to approximately 9 mg. Unfortunately, damage distribution is not always densely distributed on 25 percent of a single tooth, but is distributed across many teeth making accurate measures of material removed per tooth extremely difficult. But, this calculated value can be used as a simple rule of thumb to get a feel for the amount of debris and damage based on the gear size.

Several predictive analysis techniques were reviewed to obtain the best feature to predict damage levels from the oil debris sensor. One technique for detecting wear conditions in gear systems is by applying statistical distribution methods to particles collected from lubrication systems (Roylance (1989)). In this reference, mean particle size, variance, kurtosis, and skewness distribution characteristics were calculated from oil debris data collected off-line. The wear activity was determined by the calculated size distribution characteristics of this off-line oil debris data. In order to apply this data to on-line oil debris data, calculations were made for each reading number for each bin (Table 3.4) using the average particle size and the number of particles for each of the sixteen bins. Mean particle size, relative kurtosis, and relative skewness were calculated for each reading for 6 of the experiments with pitting damage. Appendix B discusses the statistical distribution of wear debris and shows plots of this data for experiment 2 and 3. It was not possible, however, to extract a consistent feature that increased in value from the data for all experiments. This may be due to the random nonlinear distribution of the damage progression across all 56 teeth. For this reason a more intelligent feature extraction system, applied to the oil debris mass measured by the oil debris sensor, was analyzed and will be discussed in the following paragraphs.

When defining an intelligent feature extraction system, the gear states that one plans to predict must be defined. Due to the overlap of the accumulated mass features, three primary states of the gears were identified: O.K (no gear damage); Inspect (initial pitting); Shutdown due to Damage (destructive pitting). The data from Table 3.5 was plotted in Figure 3.22. Each plot is labeled with experiment number 1 to 6. The triangles on each plot identify when the inspections were performed. The encircled triangles indicate the reading number when destructive pitting was first observed. The background color indicates the O.K., Inspect state, and Damage states. The overlap between the states is also identified with a different background color.

The changes in state for each color range were defined based on data shown in Tables 3.5 to 3.7 The minimum debris measured during experiments 1 to 6 when destructive pitting was first observed was 8.69 mg during experiment 3 and the maximum was 15.475 mg during experiment 1. The minimum and maximum debris measured during experiments 1 to 6 when destructive pitting was first observed were used to define the upper limit of the inspect scale and the lower limit of the damage scale. The maximum amount of debris measured when no damage occurred (experiment 10) was above the minimum amount of debris measured when initial pitting occurred (experiment 7). This was used as the lower limit of the inspect state. The next largest mass measured when no damage occurred (experiment 13) was used as the upper limit of the O.K. scale.

Fuzzy logic was used to extract an intelligent feature from the accumulated mass measured by the oil debris sensor. Refer to section 3.2.2 for a detailed description of fuzzy logic analysis and terminology. Membership values based on the accumulated mass measured by the oil debris sensor were identified in Figure 3.23. Membership values are defined for the 3 levels of damage: damage low (DL), damage medium (DM), and damage high (DH). Using the Mean of the Maximum (MOM) fuzzy logic defuzzification method, the oil debris mass measured during the 6 experiments with pitting damage was input into a simple fuzzy logic model created using commercially available software (Fuzzy Logic Toolbox (1998). The output of this model is shown on the Y-axis of Figure 3.24. The background color indicates the O.K., Inspect, and Damage states. From this simple fuzzy logic model, threshold limits for the accumulated mass are identified for future tests in the Spur Gear Fatigue Test Rig. Results indicate accumulated mass combined with fuzzy logic analysis techniques is an effective predictor of pitting damage on spur gears (Dempsey (2001)).

This approach has several benefits over using the accumulated mass and an arbitrary threshold limit for determining if damage has occurred. One advantage is that it eliminates the need for an expert diagnostician to analyze and interpret the data, since the output would be one of three states, O.K., Inspect, and Shutdown due to Damage. Since benign debris may be introduced into the system, due to periodic inspections, setting the lower limit to above this debris level will minimize false alarms. In addition to this, a

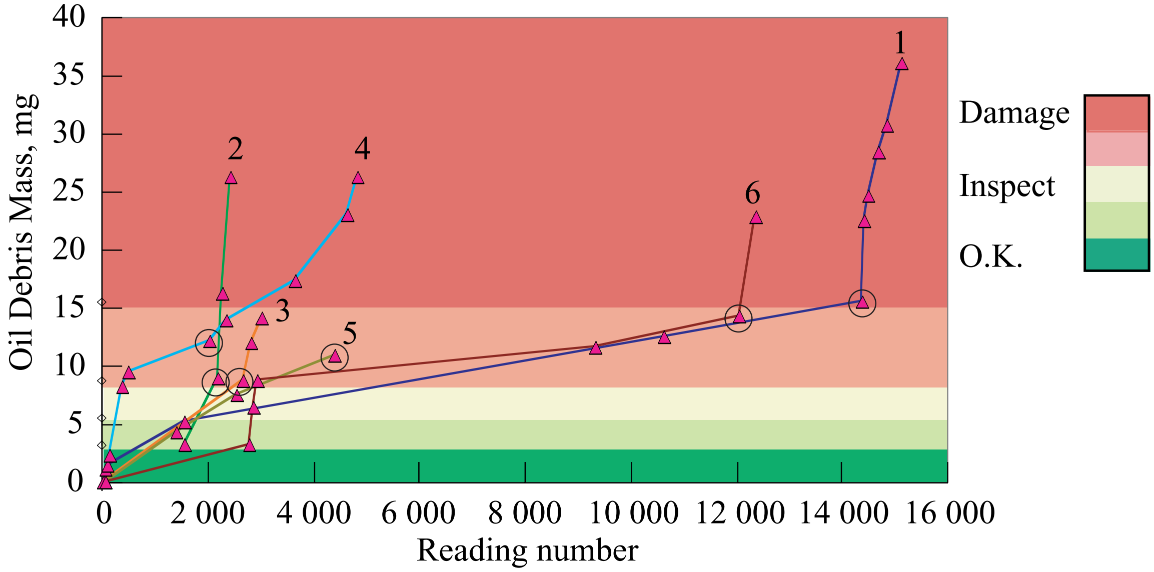


Figure 3.22.—Oil debris mass at different damage levels.

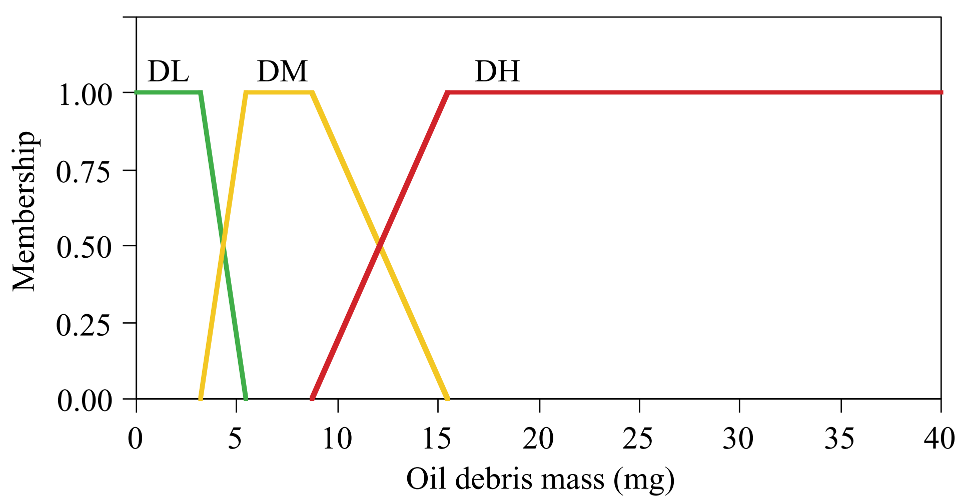


Figure 3.23.—Membership values for levels of damage.

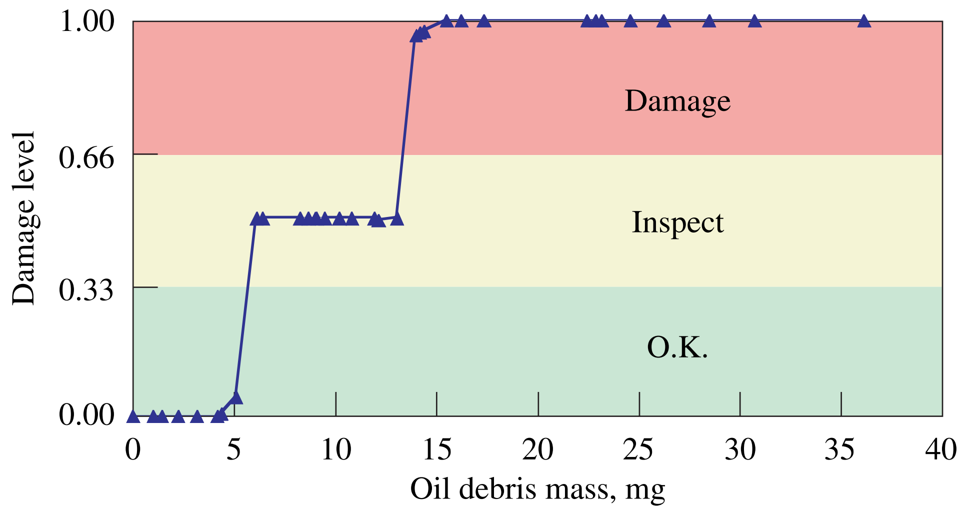


Figure 3.24.—Output of fuzzy logic model.

more advanced system can be designed with logic built-in to minimize these operational effects.

3.2 Data Analysis

3.2.1 Data Fusion Analysis

Multisensor data fusion is the process chosen to integrate oil debris and vibration based gear damage detection techniques into an intelligent health monitoring system. Multisensor data fusion is how humans use multiple senses (seeing, hearing, etc.) to make decisions about their surroundings. The International Society of Information Fusion defines information fusion as, "the theory, techniques and tools conceived and employed for exploiting the synergy in the information acquired from multiple sources such that the resulting decision or action is in some sense better than that would be possible if any of these sources were used individually" (Dasarathy (2001)). Data fusion methodology was the logical choice for integrating vibration and oil based measurement technologies for intelligent machine health monitoring

Since in many applications, the information to fuse comes from different sources, the data fusion process includes many disciplines including signal processing, statistics, artificial intelligence, cognitive psychology, and information theory. Many diagnostic tools exist to detect damage to gears. Individual tools have strengths and weaknesses for detecting damage in different environments. Combining these strengths has the potential to improve the reliability of the monitoring system. When good sensor information is used, combining multiple sensors to make decisions produces improved detection capabilities, decreased ambiguity, and increased probability an event is detected (Hall and Llinas (1997); and Hall (1992, 1999b)).

Although data fusion techniques have been used extensively in the military for target tracking and automated identification of targets, application of data fusion to condition-based maintenance has been limited to the past few years. One research paper applied this technique to vibration data collected from eight accelerometers installed on a Chinook CH-46 helicopter for fault classification (Erdley and Hall (1998)). Results of this initial research indicate an improvement when using multisensor decision level fusion techniques over single sensor processing.

There are several benefits of using sensor fusion instead of single sensor limits. Most notably, a more robust operational performance and extended spatial/temporal coverage is possible since one sensor can contribute information while others are unavailable or lack coverage of the event. For example, in a geared system, although the oil debris sensor is a good predictor of pitting damage, its ability to detect a cracked tooth has not been verified, where vibration sensors often detect cracked teeth effectively. Another benefit is increased confidence because more than a single sensor can confirm the same target or event thereby increasing assurance of its detection (Waltz (1986)).

Automated data fusion processes are used to aid human decision making by refining and reducing the quantity of information the systems operators need to examine in order to achieve timely, robust and relevant assessment of the situation. Data are combined or fused to refine state estimates and predictions.

Data fusion levels are convenient methods to categorize data fusion functions. How a specific process is performed is dependent on individual systems requirements. There is a large amount of flexibility in the approach the analyst selects including hybrid or adaptive user-defined data fusion approaches (Steinberg, et al. (1999)). Flexibility also exists when defining where in the process to fuse the information.

Sensor data can be fused at the raw data level, feature level, or decision level (Garga and Hall (1999b)). These three choices for fusing multisensor data for this application will be discussed. The first choice is direct fusion of raw sensor data. Fusion at the raw data level was not selected for this application because different types of sensors were used that required different formats and sampling rates. The second choice, feature level fusion, is to represent the sensor data into features, then to fuse these features. The features are combined into a single parameter. Faults are identified by observing changes in the signature of this parameter. Although features were obtained individually for the vibration and oil debris measurement technologies, FM4, NA4, and oil debris mass, feature level fusion was not chosen due to several reasons. Feature level fusion is best applied to the same types of measurement technologies. Decisions are made based on classification of a priori knowledge. Combining vibration and oil debris at the feature level limits the flexibility of this analysis to these 3 features. It does not provide the flexibility of parallel processing other features in the fusion process. Decision level fusion processes each sensor to achieve decisions, then combines the decisions. The third choice, decision level fusion was chosen to integrate these features. The input data is fused at the decision level. Decision level fusion was chosen because this does not limit the fusion process to a specific feature. By performing fusion at the decision level, new features can be added to the system or different features can be used without changing the entire analysis. This allows the most flexibility when applying this process to other condition based systems since, in most cases, different sensors and post-processing methods are used.

One of the reasons the application of data fusion methodology had not been widely applied to different applications was due to the lack of consistent data fusion terms. To remedy this, the U.S. Department of Defense (DoD) Joint Directors of Laboratories (JDL) Data Fusion Working Group, established in 1986, developed a data fusion model to categorize data fusion related functions (Steinberg, et al. (1999)). The model was developed to provide a conceptual framework for the functions required for a data fusion process. Their model defines a process to aid in creating a data fusion system (Garga and Hall (1999a)). While originally developed for surveillance applications, multisensor data fusion techniques can be equally applied to health monitoring as well. The architecture developed by the JDL was revised in 1998 and is shown in Figure 3.25 (Bowman (2001)). The data fusion model shows differences between fusion levels, from the source signal level to refinement levels.

Instead of discussing a generic data fusion model, the data fusion process model as applied to health monitoring will be discussed. The data fusion process as applied to health monitoring consists of several important elements (Kotanchek (1995)). The first, sources of information, relates to the accelerometers and oil debris sensor. The second, human computer interaction, relates to the expert information and system states output by computer. Sub-object assessment, refers to selecting the most relevant data for the model. Level 1 (object assessment), is the processing of data to identify precursors to fault conditions. This includes transforming the data to features and states of the geared system. Level 2 (situation assessment) is using automated reasoning to develop causal inferences such as adjusting the model for operational effects. Level 3 (impact assessment) is the development and evaluation of an alternative hypothesis regarding

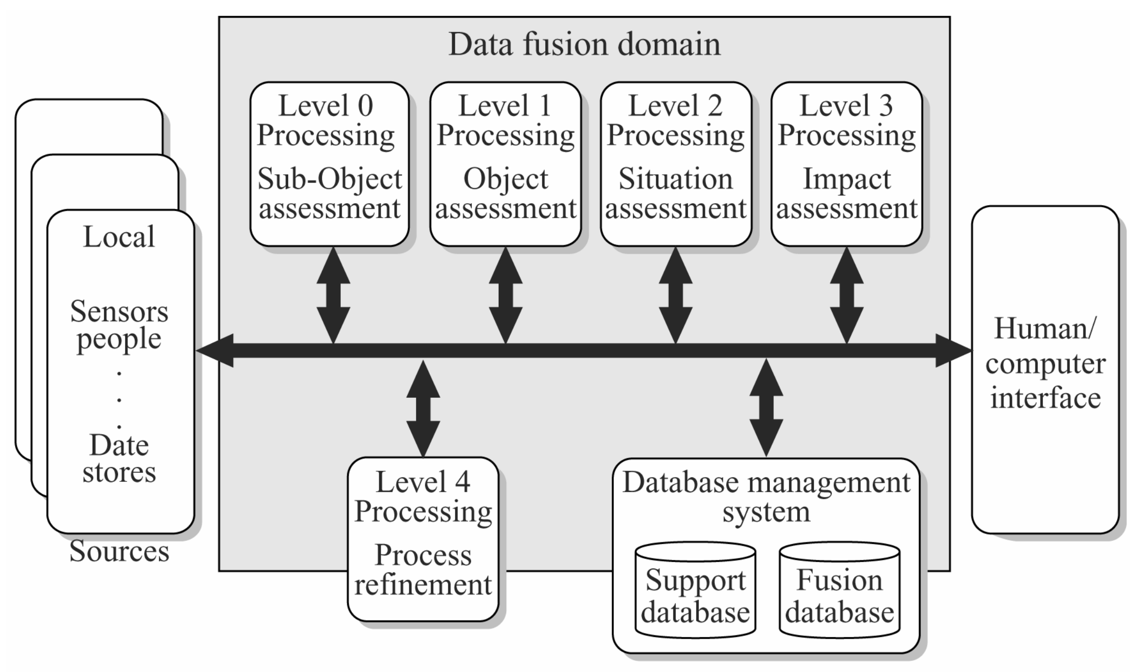


Figure 3.25.—JDL data fusion revised model (Bowman (2001)).

future events. Level 4 (process refinement) is the monitoring of the data fusion process for improvements. An example of the data fusion model applied to damage assessment and condition based maintenance was shown in the literature and is reproduced in Table 3.8 (Garga and Hall (1999a)). This thesis will focus on Level 1 elements since this is the most mature area of the data fusion process.

The data fusion model used for this application is shown in Figure 3.26. Vibration algorithms FM4 and NA4 Reset are extracted from the two accelerometers installed on the gearbox housing and through the shaft as previously shown in Figure 2.3. FM4 and NA4 Reset were calculated for each sensor. One important step in the data fusion process is the preprocessing of the sensor data. This may require the reducing the quantity of data and improving the quality of data prior to and during the feature extraction stage. For this reason, maximum values of FM4 and NA4 Reset for the two accelerometers were used as the features to input into the fuzzy logic system. The maximum values for FM4 and NA4 Reset were chosen after looking at the level of damage and the values of FM4 and NA4 Reset using the minimum, the maximum, or the mean. The maximum values of FM4 and NA4 Reset for experiments with and without damage had the least amount of variation. The accumulated mass measured is used as the feature for the oil debris sensor and will be discussed in section 3.1.4, Oil Debris Feature.

Fuzzy logic membership functions are defined for the identity declaration step. The fuzzy logic membership functions identify the damage level on each feature. The association step is used to verify the features are related. It is useful for complex systems with numerous different sensors positioned at several locations. For this simple application, association was implicit. Decision level fusion is then performed integrating membership functions with fuzzy logic rules. The output is the state of the gear. Fuzzy logic rules and membership functions applied to decision level fusion using a fuzzy logic model are discussed in the following section.

Although data fusion has been successfully applied to many applications, no consistent approach currently exists to select data fusion techniques. Work is currently underway to develop a generic framework for data fusion problem definition, classification, and an application route map (Hannah, et al. (2000)). When this work is complete, it may provide alternatives to the data fusion approach chosen for this application. Until then, selection is made based on application and user experience.

TABLE 3.8: Data fusion model applied to condition-based maintenance.
(Garga and Hall (1999a))

Rotorcraft Application	Sensors	Level 1	Level 2	Level 3
Damage Assessment	Accelerometer, temperature, pressure, strain	State of sub system; location of damaged component	Assess overall damage sustained	Predict effect of damage on continued mission and operational capabilities
Condition Based Maintenance	Accelerometer, temperature, pressure, oil debris	State of system; location of fault condition	Operations mode, damage detected.	Predict time to failure of critical components under ops demands and impact on overall Rotorcraft

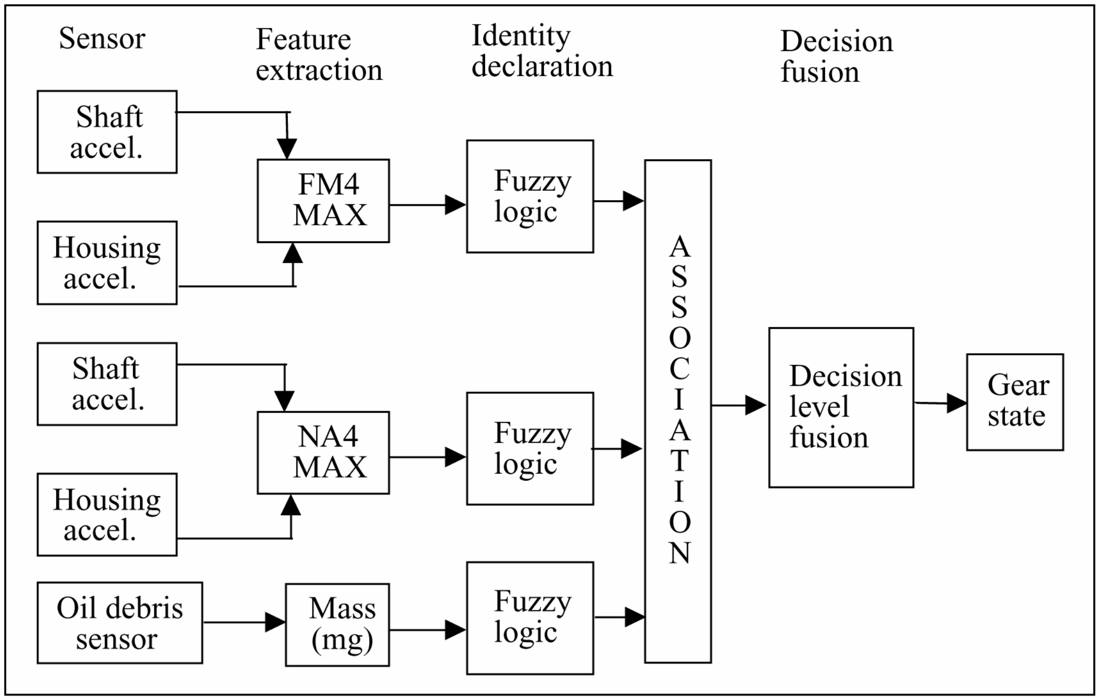


Figure 3.26.—Decision Level Fusion Model.

3.2.2 Fuzzy Logic Analysis

Fuzzy logic is used to identify the damage level on each feature and to perform the decision level fusion process on the features. Although other data fusion techniques are available, fuzzy logic was chosen based on the results of several studies to compare production rules, fuzzy logic and neural nets. One study found fuzzy logic the most robust when monitoring transitional failure data on a gearbox (Hall, et al. (1999a)). Another study comparing automated reasoning techniques for condition-based maintenance found fuzzy logic more flexible than standard logic by making allowances for unanticipated behavior (McGonigal (1997)).

Statistically based algorithms were not chosen for decision level fusion because they require large amounts of data to obtain probability inputs to the fusion system. Statistical analysis techniques use a priori knowledge about observations to make inferences (Luo, et al. (1999)). Oil debris data for the NASA Glenn Spur Gear Fatigue Test Rig did not exist outside the scope of this thesis. The lack of adequate oil debris data and the vague knowledge about integration of FM4, NA4 and the oil debris data, made it impossible to translate into a probability distribution. Statistical inference requires previous likelihood estimates and additional evidence (observations) to determine the likelihood of a hypothesis (Hall (1992)). The complexity of the data due to multiple hypothesis and multiple conditional dependent events (i.e., different levels of damage indicated by each sensor and the resulting different states of the system) made it difficult to define levels of probability for each scenario. The use of Bayesian statistical inference for this application is discussed further in Appendix D.

Fuzzy logic is a system that provides a model for human reasoning that is not exact. Fuzzy logic is precise thinking about imprecise sets, used to translate vague knowledge into a rule-based system (McGonigal (1997)). Some examples of fuzzy sets include tall people, old people, fast cars, and slow computers. Fuzzy logic applies fuzzy set theory to data, where fuzzy set theory is a theory of classes with unsharp boundaries. The data belongs in a fuzzy set based on its degree of membership (Zadeh (1992)).

Fuzzy logic was chosen for this application because it is tolerant of imprecise data, is flexible, can model nonlinear functions of arbitrary complexity, and can incorporate the experience of experts. Fuzzy logic starts with a fuzzy set, extending boolean set theory to a continuous valued logic via the concept of membership functions valued between 0 and 1. In fuzzy logic, the truth of any statement becomes a matter of degree. Fuzzy logic quantifies the extent that an attribute is imprecise, for example tall and short as compared to a precise measurement of height. Where probability quantifies the extent to which a precise concept, height, is unknown, fuzzy sets quantify fuzziness or imprecise concepts and are used to characterize nonprobabilistic uncertainties (Jang and Sun (1995); and Gibson, et al. (1994)). Construction of a fuzzy set depends on identification of a suitable universe of discourse (input space), specification of an appropriate membership function, and generation of fuzzy rules (Feng (2000)).

A membership function is a curve that defines how each point in the input space is mapped to a membership value or degree of membership between 0 and 1. The input space is sometimes referred to as the universe of discourse. The universe of discourse refers to all elements of a fuzzy set that come into consideration. The universe depends on the context. Every element in the universe of discourse is a member of a fuzzy set to some degree. The only condition a membership function must satisfy is that it is a continuous function that varies between 0 and 1. When X is equal to the universe of discourse and x are its elements, fuzzy set A in X is defined as a set of ordered pairs:

$$A = \{x, \mu_A(X) \mid x \in X\} \tag{3.6}$$

where, $\mu_A(x)$ is the membership function of x in A. Each element of X is mapped to a membership function value between 0 and 1.

Fuzzy sets are sets with degrees of membership with gradual transitions from membership to nonmembership and the degree of membership is a real number from 0 to 1. How does one determine the type of membership function to use for a specific application? Deciding the type of membership function to use for a specific application is subjective and nonrandom as compared to probability theory that deals with objective treatment of random phenomena. A few rules of thumb have been listed in the literature (Jantzen (1999a) and Jantzen (1999b)). One is that the width of the membership functions must be wide enough to allow for noise in the measurement. For example, Figure 3.27 shows a trapezoidal membership function for two states. The x-axis refers to the feature values and the y-axis refers to the degree of membership. Fuzzification looks up the degree of membership for the feature value. The feature value must be between the minimum (0) and maximum (9) values for a membership function to be defined. If the feature value is less than 0, or greater than 9, a membership value is not defined (Feng, et al. (2000)). Another rule of thumb is the states require a certain amount of overlap. If a gap exists between the states, no rules will fire for values in the gap and the membership

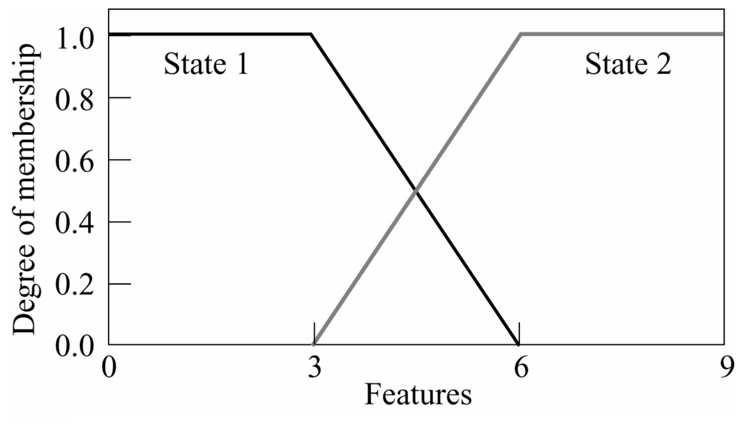


Figure 3.27.—Fuzzy logic membership functions.

function output is undefined. Referring to Figure 3.27, state 1 and state 2 overlap by a value of 3.

Although membership functions can take many shapes, one recommendation for developing new membership functions is to start with triangular shaped membership functions (Jantzen (1999b)). Triangular and trapezoidal membership functions have the advantage of simplicity, they are formed using straight lines instead of complicated equations. The simple formulas required for triangular and trapezoidal membership functions and their computational efficiency make them suitable for real-time implementation. One benefit of using simple triangular and trapezoidal membership functions is the ease of modifying the membership functions for other geared application. Work has been done in designing a fuzzy logic system where the rules and membership functions remain the same for different systems. The only parameters that need to be changed are the input/output scaling factors on the membership functions (Yeh (1999)). This approach appeared to be a feasible method for applying membership functions developed for the Spur Gear Fatigue Rig to other fatigue rigs by rescaling the membership functions. For this reason, triangular and trapezoidal membership functions were chosen for this analysis and their specific application will be discussed in the following chapters.

Once membership functions are defined for a fuzzy set, fuzzy rules must be defined. Fuzzy rules are defined by experts in the field. Experts express their field knowledge in rules with an IF-THEN format. The developer of vibration diagnostic tool NA4 had over 12 years of experience using FM4 in the NASA Glenn fatigue studies. Fuzzy rule development was therefore conducted based on the “expert” input from this source. At the time of this work, a database of oil debris damage data for the oil debris sensor did not exist. The experience gained in this study with the oil debris sensor was used for development of fuzzy rules. An example of a fuzzy if-then rule is, “if x is A then y is B.” A and B are linguistic values defined by fuzzy sets on universes of discourse X and Y. Often “x is A” is called the antecedent or premise and “y is B” is called the consequence or conclusion (Feng, et al. (2000)). A and B are fuzzy sets on a mutual universe. The intersection of A and B is defined as:

$$A \cap B = AND = a \min b \tag{3.7}$$

The union of A and B is defined as:

$$A \cup B = OR = a \max b \tag{3.8}$$

Fuzzy inference, the process of formulating the mapping from a given input to an output using fuzzy logic, provides a basis from which decisions are made. The fuzzy system is defined by 3 main components: (1) fuzzy input/output variables defined by their fuzzy variables; (2) a set of fuzzy rules; and (3) fuzzy inference mechanism (Kasabov (1996)). Mamdani's fuzzy inference system is the most commonly used fuzzy methodology (Mamdani and Assilan (1975)). It is based on Zadeh's pioneering paper on fuzzy algorithms for decision processes (Zadeh (1973)). In the Mamdani type inference systems the output membership functions are fuzzy sets. The process is detailed below and in Figure 3.28:

1. Fuzzify inputs or fuzzification: converts each piece of input data to degrees of membership by a lookup in one of several membership functions.
2. Apply fuzzy operator: AND = minimum; OR = maximum
3. Apply implication methods: apply weight to rule; output fuzzy set is truncated and scaled.
4. Aggregate all outputs: aggregation is the process by which fuzzy sets represent the outputs of each rule and are combined into a single fuzzy set.
5. Defuzzify: calculate a single output value from a fuzzy set. Convert fuzzy membership information into a crisp output.

The process described above will be used to define the fuzzy logic model for health monitoring of gears. The inputs are the damage detection features, the rules are defined based on the expertise of the diagnostician, and the outputs are the states of the gears. Several defuzzification methods exist. There is no systematic procedure for choosing a defuzzification strategy. The Mean of the Maximum (MOM) defuzzification method was chosen because it gave the most plausible results for this application. The MOM method finds the output with the maximum membership and takes the x-axis average of all points with this maximum membership value. If there is more than one point that has maximum degree, the mean of the points are taken. An example for this model will be discussed in section 3.3, Feature Validation for Sensor Fusion.

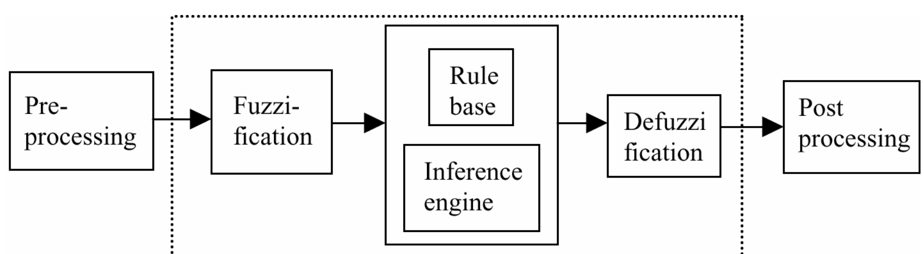


Figure 3.28.—Mamdani's fuzzy inference system. (Jantzen (1999b)).

Commercially available software was used to build the model because it provided a convenient method to map an input space to an output space (Fuzzy Logic Toolbox (1998)). The chosen software allows the user to create and edit fuzzy inference systems. The input space consists of level of damage indicated by the oil debris and vibration features oil debris mass, FM4, and NA4 Reset. Membership functions for the features are defined as levels of damage. Levels of damage are damage low (DL), damage medium (DM), and damage high (DH).

The levels of damage for each feature are as follows: oil debris mass (DL, DM, DH), NA4 Reset (DH, DL), and FM4 (DH, DL). Output space is defined as the state of the gear. The three states of the gear are O.K. (no gear damage), inspect (initial pitting), and shutdown due to damage (severe destructive pitting). The membership functions for each feature will be discussed in detail in the Chapter 4, subchapter 4.1, Assessment of Diagnostic Features Integration.

3.3 Feature Validation for Sensor Fusion

The analysis discussed in this section is based on data collected during 24 experiments, 15 of which had pitting damage occur. Video inspection images are available for 13 of the experiments with pitting damage, 2 were performed prior to installation of the video inspection system.

Table 3.9 is a summary of the experiments performed and a description of the damage. The second column lists the reading number the pitting was first observed via video or manual inspection. Video inspection was used during Experiments 1 to 6 and 18 to 24. Manual inspection was used for experiments 7 to 17. The "oil debris" column is the amount of debris measured at this reading. The last reading collected for this experiment is listed in the fifth column. All gears were visually inspected at test completion and the damage description and amount of debris at this time are listed in the last 2 columns. The damage description gives the damage observed on the driver (Dr) and driven (Dn) gears. Damage is defined as initial pitting (ip), and destructive pitting (de) to the total number of teeth for each gear. For example, Dr: de 3t, ip allt, is driver gear had destructive pitting on 3 teeth and initial pitting on all of the teeth. A detailed description of the damage to each tooth was correlated with the video images for each experiment. Detailed damage descriptions for the experiments with damage (experiments 1 to 8 and 18 to 24) are shown in Tables 3.10 to 3.24. The damage progression images of one damaged tooth for experiments 1 to 8 and 18 to 24 are shown in Figures 3.6, 3.9, 3.12 to 3.14, and 3.29 to 3.38. The table listing the amount of debris at test completion for experiments 9 to 17 with no damage was shown in section 3.1.4, Table 3.7. The damaged tooth chosen for each experiment to display in the figures is identified in bold in the damage description tables. As mentioned previously, damage is shown on less than half of the tooth because the test gears are run offset to provide a narrow effective face width to maximize gear contact stress. As can be seen from the tables and images, the amount of damage to the gear teeth and the mass of debris measured varied significantly for each experiment. The video inspection system was used to minimize subjective observations of the damage to the tooth.

TABLE 3.9
Summary of Experiments

Experiment Number	Rdg Pitting 1^{st} Observed	Damage Description	Oil Debris (mg)	Rdg at Test Completion	Damage Description	Oil Debris (mg)
1	14369	Dr: de 1t Dn: de 1t	15.475	15136	Dr: de 3t, ip allt Dn: de3t	36.108
2	2199	Dr: Dn: de 1t	8.934	2444	Dr: de 2t, ip allt Dn de 3t, ip 3t	26.268
3	2669	Dr: de 2t Dn: de 2t	8.690	3029	Dr: de 3t, ip allt Dn: de 3t, ip3t	14.148
4	2065	Dr: de 3t Dn:	12.132	4863	Dr: de 7t, ip allt Dn de 3t, ip allt	26.227
5	2566	Dr: ip 2t Dn:	7.413	4425	Dr: de 11t, ip allt Dn de 10t, ip allt	10.811
6	12061	Dr: Dn: de 1t	14.365	12368	Dr: de 1t, ip allt Dn de 2t, ip allt	22.851
7				13716	Dr: ip 1t Dn ip 1t	3.381
8	5181	Dr: ip 2t Dn: ip 3t	6.012	5314	Dr: de 6t, ip8t Dn de6t, ip7t	19.101
9				29866	No damage	2.359
10				20452	No damage	5.453
11				204	No damage	.418
12				15654	No damage	2.276
13				25259	No damage	3.159
14				5322	No damage	0
15				21016	No damage	.125
16				380	No damage	.099
17				21066	No damage	.064
18				888	Dr: de 6t, ip allt Dn de 4t, ip allt	22.541
19				199	Dr: de 3t, ip allt Dn de 1t, ip allt	11.230
20				1593	Dr: de 1t, ip allt Dn ip allt	5.346
21	317	Dr: de 1t Dn: de 1t	4.04	514	Dr: de 2t, ip allt Dn: de 2t, ip allt	17.912
22				838	Dr: ip 5t Dn: de 3t, ip allt	7.224
23				10688	Dr: de 2t, ip allt Dn: de 1t, ip allt	6.399
24	7170	Dr: de 1t Dn:	6.186	7224	Dr: de 1t, ip allt Dn: ip allt	9.681

TABLE 3.10
Damage Description for Experiment 1

Reading Number Run Time (min)	Damage Description	Teeth damaged on Driver Gear	Teeth damaged on Driven Gear	Oil Debris Mass (mg)
60	Run-in Wear	All	All	1.003
120	Wear	All	All	1.418
1581	Wear	All	All	5.113
10622	Wear	All	All	12.533
14369	Wear Destructive Pitting	All 6	All 6	15.475
14430	Wear Destructive Pitting	All 6	All 6	22.468
14512	Wear Destructive Pitting	All 6,7	All 6,7	24.586
14688	Wear Destructive Pitting	All 6,7	All 6,7	28.451
14846	Wear Destructive Pitting	All 6,7	All 6,7	30.686
15136	Wear Initial Pitting Destructive Pitting	All teeth All teeth **6**,7,8	All teeth **6**,7,10	36.108

TABLE 3.11
Damage Description for Experiment 2

Reading Number Run Time (min)	Damage Description	Teeth Damaged on Driver Gear	Teeth Damaged on Driven Gear	Oil Debris Mass (mg)
1573	Run-in Wear	All	All	3.285
2199	Wear Destructive Pitting	All	All 11	8.934
2296	Wear Destructive Pitting	All	All 10, 11	16.267
2444	Wear Initial Pitting Destructive Pitting	All All 10,**11**	All 10, 11, 14 10,**11**,14	26.268

TABLE 3.12
Damage Description for Experiment 3

Reading Number Run Time (min)	Damage Description	Teeth Damaged on Driver Gear	Teeth Damaged on Driven Gear	Oil Debris Mass (mg)
58	Run-in Wear	All	All	0
2669	Wear Destructive Pitting	All 1,28	All 1,28	8.69
2857	Wear Destructive Pitting	All 1,6,28	All 1,6,28	11.889
3029	Wear Initial Pitting Destructive Pitting	All All 1,6,**28**	All 1,6,28 1,6,**28**	14.148

TABLE 3.13
Damage Description for Experiment 4

Reading Number Run Time (min)	Damage Description	Teeth damaged on Driver Gear	Teeth damaged on Driven Gear	Oil Debris Mass (mg)
64	Run-in Wear	All	All	0
150	Wear	All	All	2.233
378	Wear	All	All	8.297
518	Wear	All	All	9.462
2065	Wear Destructive Pitting	All 3,6,7	All	12.132
2366	Wear Destructive Pitting	All 3,6,7	All	13.977
3671	Wear Destructive Pitting	All 3,6,7	All 5	17.361
4655	Wear Destructive Pitting	All 3,6,7	All 3,5,7	23.12
4863	Wear Initial Pitting Destructive Pitting	All All 3,6,**7**,11,14,16,28	All All 3,5,**7**	26.227

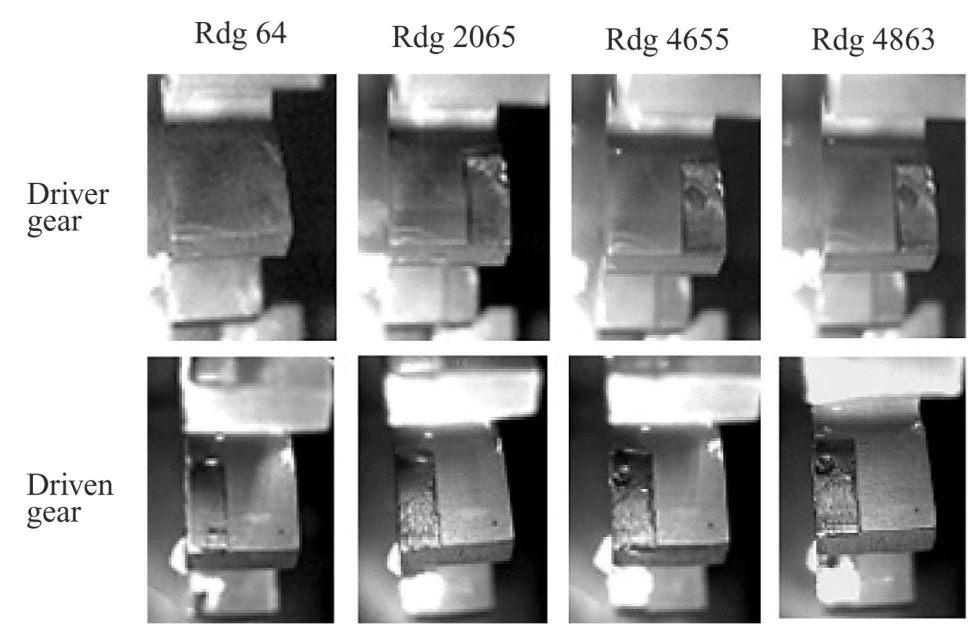


Figure 3.29.—Damage progression of driver/driven tooth 7 for experiment 4

TABLE 3.14
Damage Description for Experiment 5

Reading Number Run Time (min)	Damage Description	Teeth Damaged on Driver Gear	Teeth Damaged on Driven Gear	Oil Debris Mass (mg)
62	Run-in Wear	All	All	0
1405	Wear	All	All	4.214
2566	Wear Destructive Pitting	All 17,25	All	7.413
4425	Wear Initial Pitting Destructive Pitting	All All 1,3,17,18,19,20, 21,24,25,**26**,28	All All 1,17,18,19, 20, 21,22,24,25,**26**	10.811

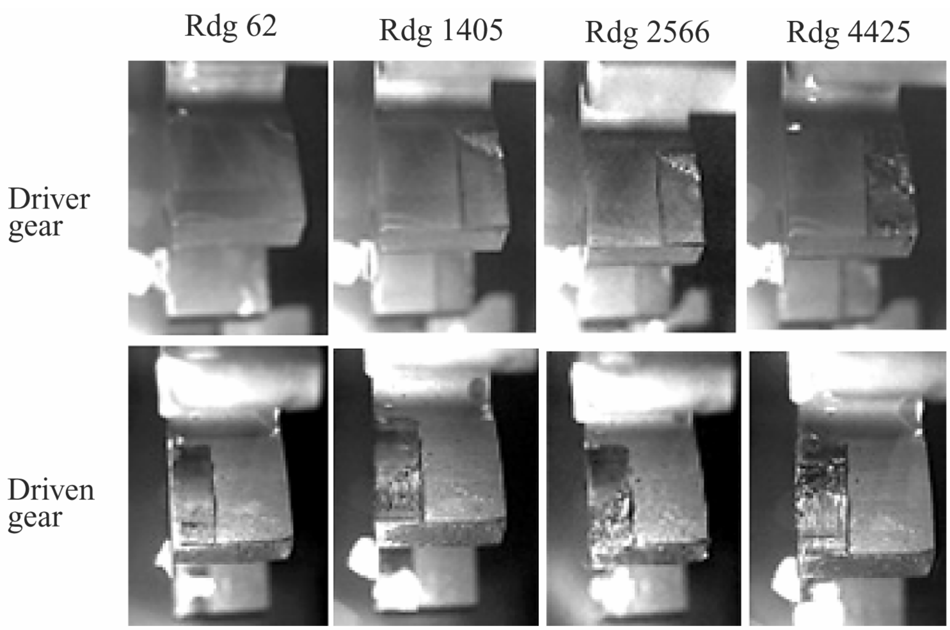


Figure 3.30.—Damage progression of driver/driven tooth 26 for experiment 5

TABLE 3.15
Damage Description for Experiment 6

Reading Number Run Time (min)	Damage Description	Teeth Damaged on Driver Gear	Teeth Damaged on Driven Gear	Oil Debris Mass (mg)
60	Run-in Wear	All	All	0
2810	Wear	All	All	3.192
2885	Wear	All	All	6.396
2957	Wear	All	All	8.704
9328	Wear	All	All	11.692
12061	Wear Destructive Pitting	All	All 22	14.365
12368	Wear Initial Pitting Destructive Pitting	All All 25	All All **22**,25	22.851

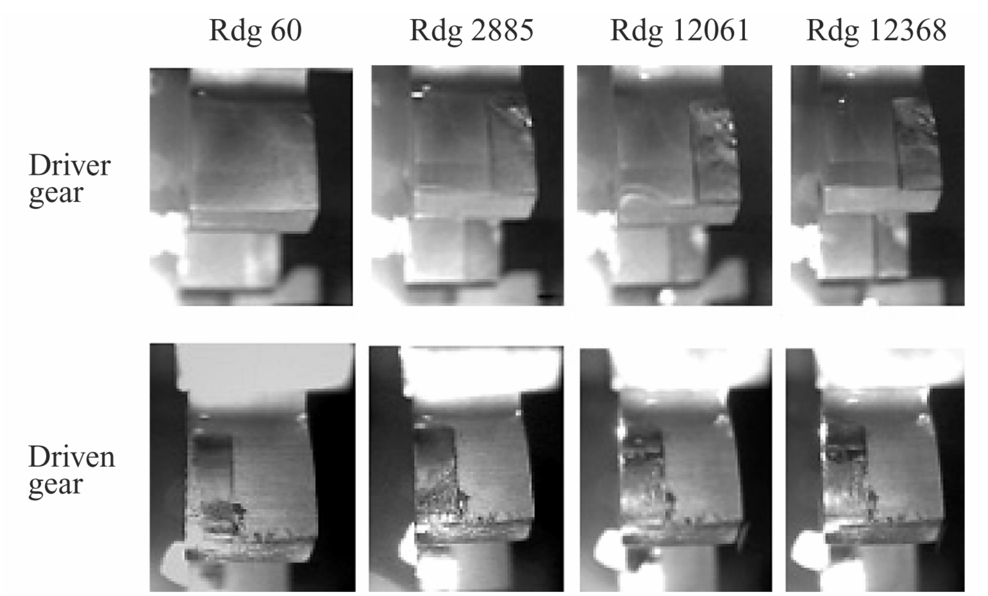


Figure 3.31.—Damage progression of driver/driven tooth 22 for experiment 6

TABLE 3.16
Damage Description for Experiment 7

Reading Number Run Time (min)	Damage Description	Teeth Damaged on Driver Gear	Teeth Damaged on Driven Gear	Oil Debris Mass (mg)
13716	Initial Pitting	12	12	3.381

TABLE 3.17
Damage Description for Experiment 8

Reading Number Run Time (min)	Damage Description	Teeth Damaged on Driver Gear	Teeth Damaged on Driven Gear	Oil Debris Mass (mg)
5181	Initial Pitting Destructive Pitting	15, 16	15,16, 17	6.012
5314	Initial Pitting Destructive Pitting	19,24, 27 9,15,16,17,18, 24	14 9,15,16,17,18, 24	19.101

TABLE 3.18
Damage Description for Experiment 18

Reading Number Run Time (min)	Damage Description	Teeth Damaged on Driver Gear	Teeth Damaged on Driven Gear	Oil Debris Mass (mg)
60	Run-in Wear	All	All	0
888	Wear Initial Pitting Destructive Pitting	All All 5,6,**7**,11,25,27	All All 5,6,**7**,11	22.541

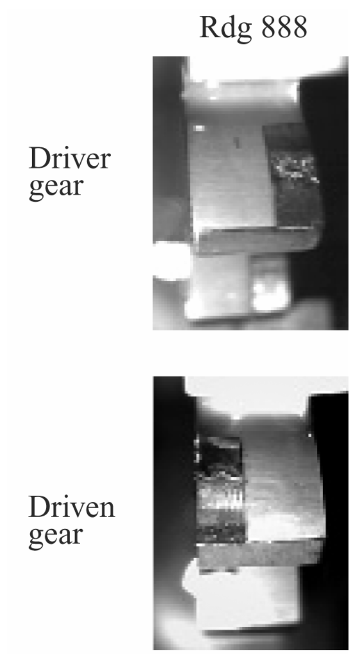


Figure 3.32.—Damage progression of driver/driven tooth 7 for experiment 18.

TABLE 3.19
Damage Description for Experiment 19

Reading Number Run Time (min)	Damage Description	Teeth Damaged on Driver Gear	Teeth Damaged on Driven Gear	Oil Debris Mass (mg)
62	Run-in Wear	All	All	0
	Wear	All	All	
	Initial Pitting	All	All	
199	Destructive Pitting	**3**, 4, 13	**3**	11.230

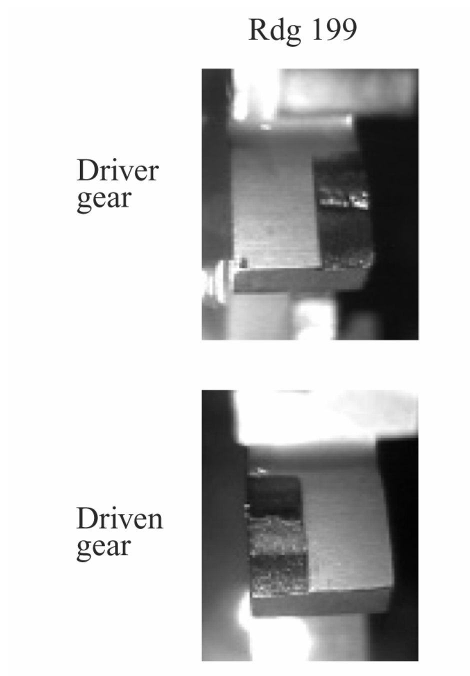


Figure 3.33.—Damage progression of driver/driven tooth 3 for experiment 19

TABLE 3.20
Damage Description for Experiment 20

Reading Number Run Time (min)	Damage Description	Teeth Damaged on Driver Gear	Teeth Damaged on Driven Gear	Oil Debris Mass (mg)
66	Run-in Wear	All	All	0
	Wear	All	All	
	Initial Pitting	All	All	
1593	Destructive Pitting	**26**		5.346

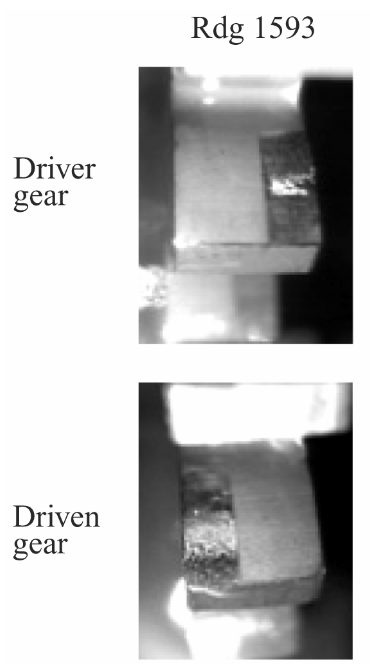


Figure 3.34.—Damage progression of driver/driven tooth 26 for experiment 20

TABLE 3.21
Damage Description for Experiment 21

Reading Number Run Time (min)	Damage Description	Teeth Damaged on Driver Gear	Teeth Damaged on Driven Gear	Oil Debris Mass (mg)
62	Run-in Wear	All	All	0
317	Wear Destructive Pitting	All 22	All 22	4.04
373	Wear Destructive Pitting	All 22,24	All 22,24	9.808
455	Wear Destructive Pitting	All 22,24	All 22,24	10.936
514	Wear Initial Pitting Destructive Pitting	All All **22**,24	All All **22**,24	17.912

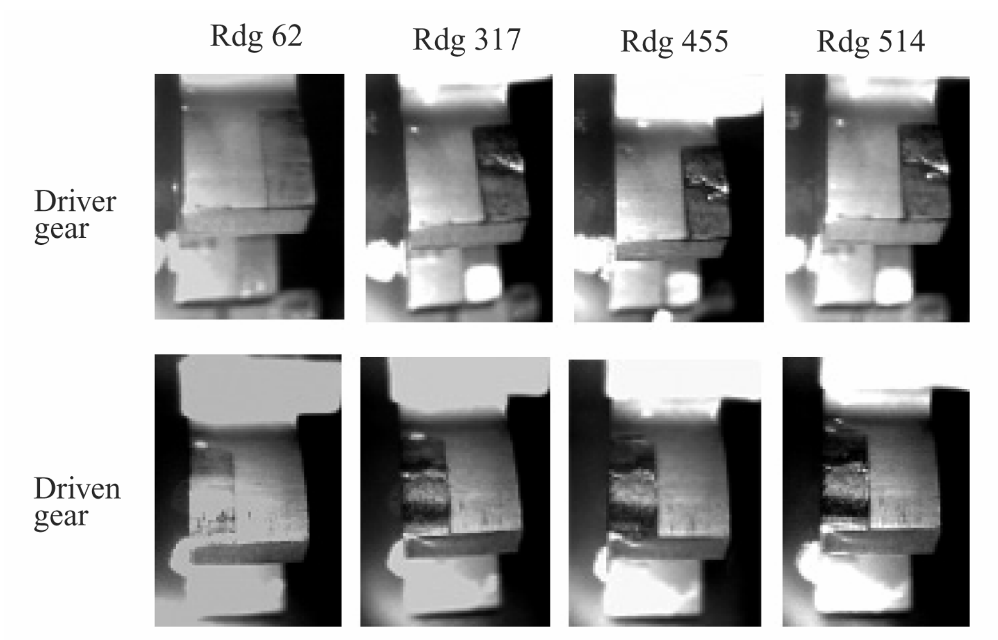


Figure 3.35.—Damage progression of driver/driven tooth 22 for experiment 21

TABLE 3.22
Damage Description for Experiment 22

Reading Number Run Time (min)	Damage Description	Teeth Damaged on Driver Gear	Teeth Damaged on Driven Gear	Oil Debris Mass (mg)
129	Run-in Wear	All	All	0
838	Wear Initial Pitting Destructive Pitting	All 12,14,**18**,19,26	All All **18**, 19, 26	7.224

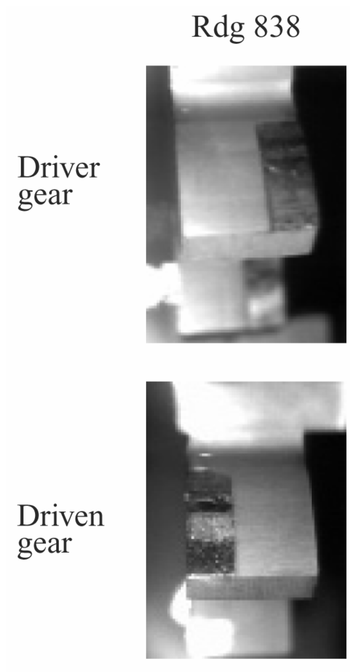


Figure 3.36.—Damage progression of driver/driven tooth 18 for experiment 22

TABLE 3.23
Damage Description for Experiment 23

Reading Number Run Time (min)	Damage Description	Teeth Damaged on Driver Gear	Teeth Damaged on Driven Gear	Oil Debris Mass (mg)
62	Run-in Wear	All	All	0
10688	Wear Initial Pitting Destructive Pitting	All All **10**,11	All All **10**	6.399

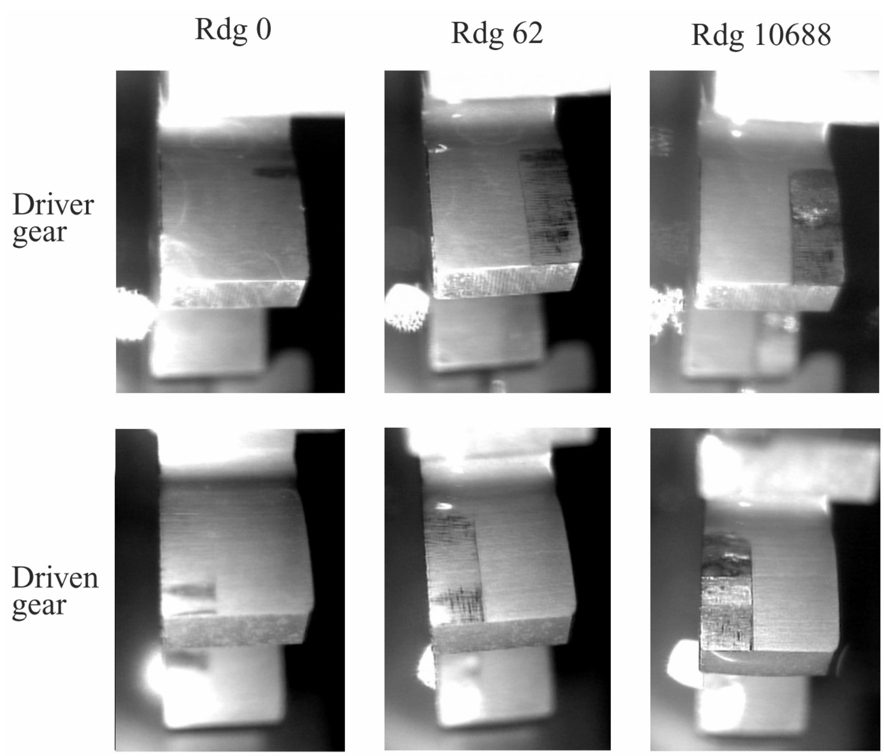


Figure 3.37.—Damage progression of driver/driven tooth 10 for experiment 23

TABLE 3.24
Damage Description for Experiment 24

Reading Number Run Time (min)	Damage Description	Teeth Damaged on Driver Gear	Teeth Damaged on Driven Gear	Oil Debris Mass (mg)
0	Run-in Wear	All	All	0
4077	Wear	All	All	2.344
7170	Wear Destructive Pitting	All 26	All	6.186
7224	Wear Initial Pitting Destructive Pitting	All All **26**	All All	9.681

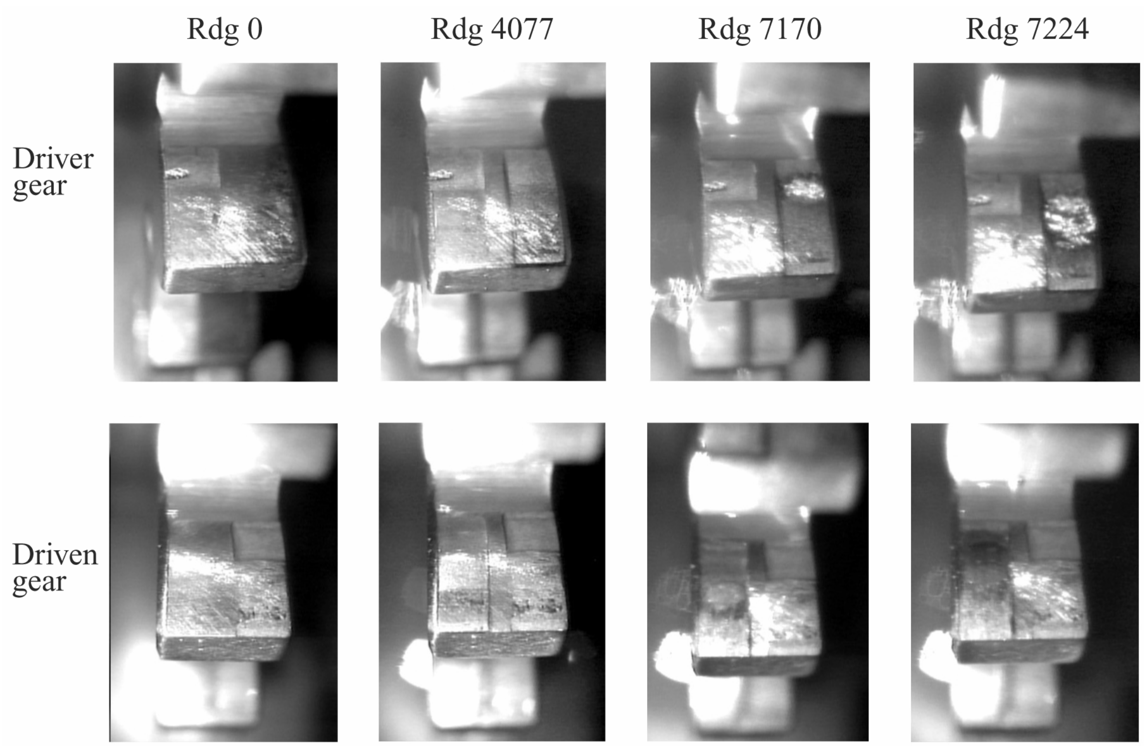


Figure 3.38.—Damage progression of driver/driven tooth 26 for experiment 24

Fuzzy logic techniques were applied to the oil debris and vibration data in order to build a simple data fusion model that predicts when pitting damage occurs on one or more teeth. The model was developed to predict three states of the gears: O.K (no gear damage); Inspect (initial/destructive pitting); Shutdown due to Damage (severe destructive pitting).

Membership values to use in the fuzzy logic model were defined for the three features: oil debris, FM4 and NA4. For the oil debris sensor, membership values were based on the accumulated mass and the amount of damage observed on the video images and by visual inspection. Membership values are defined for 3 levels of damage: damage low (DL), damage medium (DM), and damage high (DH). The process used to define membership functions for the oil debris sensor is discussed in section 3.1.4, Oil Debris Feature, and indicates accumulated mass is a good predictor of pitting damage on spur gears and fuzzy logic is a good technique for setting threshold limits that discriminate between states of pitting wear. The membership values for the oil debris mass are shown in Figure 3.23. The membership functions were defined using data collected from experiments 1 to 17.

Membership values were defined for 2 levels of damage for vibration algorithms FM4 and NA4 Reset: damage low (DL) and damage high (DH). From observations made during experiments with pitting damage, it was found that FM4 and NA4 Reset increase in magnitude initially, then decrease as damage progresses. Since damage progression could not be detected from the vibration algorithms, only two levels of damage were defined for FM4 and NA4 Reset. FM4 and NA4 Reset were calculated for both the accelerometer on the shaft and the accelerometer on the housing. It was also determined from the experimental data that although FM4 and NA4 Reset were calculated for both accelerometers, using the maximum value of the two accelerometers was more reliable than using them separately or using the average of the two. And, since the magnitude of NA4 Reset was significantly larger than FM4 when pitting damage began to occur, different membership functions were defined for each algorithm.

The maximum FM4 values for experiments when damage occurred between inspection intervals are shown in Tables 3.25 and 3.26. The maximum FM4 values for experiments when no damage occurred are listed in Table 3.27. Membership values defined for the 2 levels of damage for FM4 are shown in Figure 3.39. The damage low membership function X,Y coordinates [0, 1; 4.04, 1; 7.68, 0], and damage high membership function X,Y coordinates [4.04, 0; 7.68, 1; 10, 1], were defined by plotting and analyzing the data in Tables 3.25 and 3.26. Trial and error was used to determine the coordinates that minimized false alarms and decreased missed hits. Due to the insensitivity of FM4 to damage progression, logic was also programmed into the model to freeze FM4 when it exceeded 7.68.

The maximum NA4 Reset values for experiments when damage occurred between inspection intervals are shown in Tables 3.28 and 3.29. The maximum NA4 Reset values for experiments when no damage occurred are listed in Table 3.30. It should be

TABLE 3.25
Reading Numbers and FM4 Max at Video Gear Inspection

Experiment 1		Experiment 2		Experiment 3		Experiment 4		Experiment 5		Experiment 6	
Rdg#	FM4 Max	Rdg#	FM4 Max	Rdg#	FM4 Max	Rdg#	FM4 Max	Rdg#	FM4 Max	Rdg#	FM4 Max
60		1573	3.52	58		64		62		60	
120	3.68	2199	5.23	2669	4.66	150	3.04	1405	3.63	2810	3.78
1581	3.53	2296	5.03	2857	5.91	378	3.97	2566	3.10	2885	3.35
10622	3.75	2444	6.09	3029	3.86	518	2.94	4425	4.04	2957	3.29
14369	4.39					2065	3.07			9328	3.76
14430	8.41					2366	4.19			12061	3.66
14512	7.37					3671	2.94			12368	4.13
14688	7.21					4655	5.64				
14846	7.16					4863	5.49				
15136	7.01										

*Note: Highlighted cells identify reading and mass when destructive pitting was first observed

TABLE 3.26
FM4 max during experiments
with visual gear inspection

Experiment 7		Experiment 8	
Rdg#	FM4 Max	Rdg#	FM4 Max
13716	7.68	5314	9.90

TABLE 3.27
FM4 max at completion of
experiments with no damage

Experiment	Readings	FM4 Max
9	29866	7.36
10	20452	4.61
11	204	3.78
12	15654	4.18
13	25259	7.60
14	5322	5.05
15	21016	5.21
16	380	3.54
17	21066	5.17

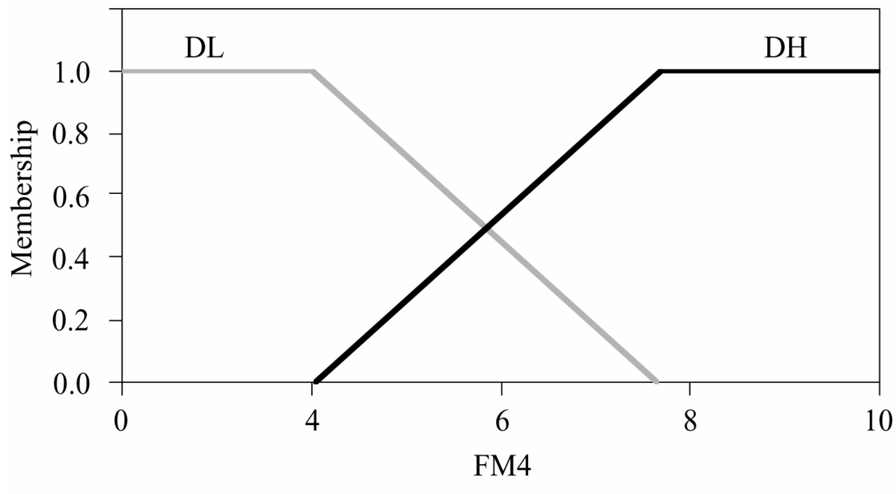


Figure 3.39.—Membership values for FM4 feature

noted that a load measurement was not available during experiments 13 to 17, so NA4 Reset corrections were not made for these experiments. Membership values defined for the 2 levels of damage for NA4 Reset are shown in Figure 3.40. The damage low membership function X,Y coordinates [0, 1; 7.36, 1; 12.6, 0], and damage high membership function X,Y coordinates [7.36, 0; 12.6, 1; 55, 1], were defined by plotting and analyzing the data in Table 3.28. Trial and error was used to determine the coordinates that minimized false alarms and decreased missed hits. Due to the insensitivity of NA4 Reset to damage progression, logic was also programmed into the model to freeze NA4 Reset when it exceeded 12.60.

Due to the findings of this experimental research, post processing was required on the data prior to inputting the data into the fuzzy logic membership functions. These modifications were performed on the experimental data to improve the integrity of the individual features prior to data fusion. A block diagram showing the preprocessing performed on the oil debris and vibration data collected during these experiments is shown on Figure 3.41.

The inputs to the fuzzy system are the membership functions discussed in the preceding paragraphs. The degree of membership for the output of the fuzzy model is shown in Figure 3.42 with the output status or state of the gear: O.K (no gear damage); Inspect (initial pitting); Shutdown due to damage (destructive pitting). The output was defined to give the end user a simple function based on the state of the gear. A value of 0 to 0.33 indicates the gear is O.K., 0.33 to 0.66 indicates the gear should be inspected, and 0.66 to 1.0 indicates shutdown the system, the gear is damaged. The rules defined for the model are listed in Table 3.31. A simple interpretation of one of the rules is as follows:

TABLE 3.28
Reading Numbers and NA4 Reset Max at Video Gear Inspection

Experiment 1		Experiment 2		Experiment 3		Experiment 4		Experiment 5		Experiment 6	
Rdg#	NA4 Max	Rdg#	NA4 Max	Rdg#	NA4 Max	Rdg#	NA4 Max	Rdg#	NA4 Max	Rdg#	NA4 Max
60		1573	13.21	58		64		62		60	
120	4.50	2199	20.76	2669	30.60	150	3.72	1405	4.18	2810	6.65
1581	4.30	2296	7.72	2857	7.16	378	4.88	2566	4.49	2885	3.57
10622	8.56	2444	7.17	3029	11.38	518	4.92	4425	7.36	2957	4.45
14369	40.55					2065	4.77			9328	5.03
14430	9.16					2366	5.82			12061	10.38
14512	9.19					3671	6.33			12368	6.96
14688	8.44					4655	12.60				
14846	10.41					4863	2.84				
15136	7.83										

*Note: Highlighted cells identify reading and mass when destructive pitting was first observed

TABLE 3.29
NA4 Reset max during experiments
with visual gear inspection

Experiment 7		Experiment 8	
Rdg#	NA4 Max	Rdg#	NA4 Max
13716	54.03	5314	11.45

TABLE 3.30
NA4 Reset max at completion of
experiments with no damage

Experiment	Readings	NA4 Max
9	29866	14.98
10	20452	11.62
11	204	4.82
12	15654	9.66
13	25259	40.9
14	5322	13.1
15	21016	40.5
16	380	7.59
17	21066	8.76

*Note: No load corrections to NA4 Reset for experiments 13 to 17.

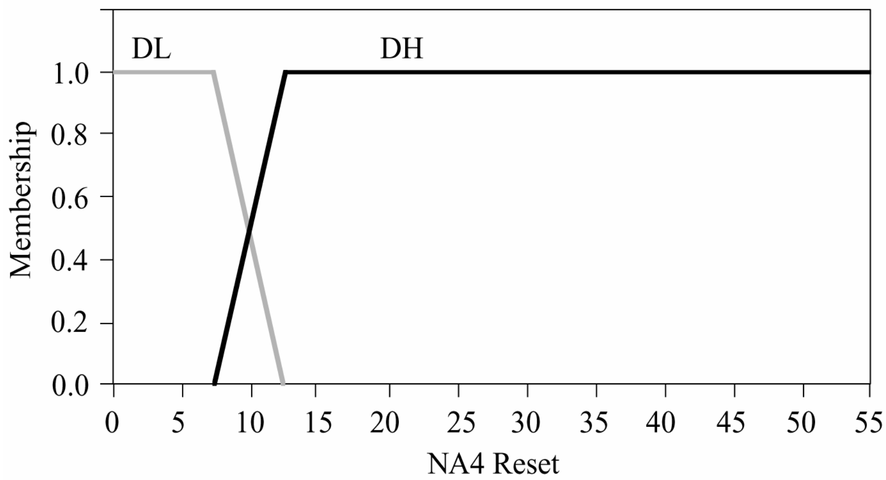


Figure 3.40.—Membership values for NA4 Reset feature

for rule 1, if FM4 indicates damage is low, NA4 indicates damage is low, and the oil debris indicates the damage is low, then the gear is O.K. Using the membership values and rules for the vibration and oil debris features, and the Mean of the Maximum (MOM) fuzzy logic defuzzification method, a simple fuzzy logic model was developed. The model was defined using the data collected from experiments 1 to 17. The input/output data to the fuzzy model for each experiment will be discussed in the following paragraphs.

An example of the membership function outputs when rules are applied or fired is shown in Figure 3.43. The numbers on the left hand side of the figure refer to the rules 1 through 12 listed in Table 3.31. The input values and membership functions for FM4, NA4, and the oil debris are listed in the first 3 columns. Based on the inputs for each feature (5.5, 27.6, and 20), the rules that are fired are shaded. The last column shows the output membership functions and the rules that are fired are shaded. For example, rule 1, "If (FM4 is DL) and (NA4 is DL) and (debris is DL) then (output is O.K)", no output membership functions are fired because the minimum (and) membership functions for the 3 parameters is zero. For this example, Rule 2 and Rule 11 fire the shutdown membership function in the output. Since the mean of the maximum defuzzification method was chosen, the maximum output membership function of rule 11 is used. The mean of maximum on the x-axis of the output is .935, shown in the last row of the output column.

Figures 3.44 to 3.47 are representative plots for 4 of the 24 experiments. Each figure is comprised of 2 plots. The plot on the top is a plot of the 3 features measured during each experiment. FM4 and NA4 Reset correspond to the left Y-axis, the accumulated mass measured by the oil debris sensor corresponds to the right Y-axis. These features are input into the model developed for this research. The plot on the bottom is the model output. The triangles on the X-axis correspond to the inspection reading numbers. The background colors in different shades indicate O.K., Inspect, and Shutdown states. A short description of Figures 3.44 to 3.47 will follow.

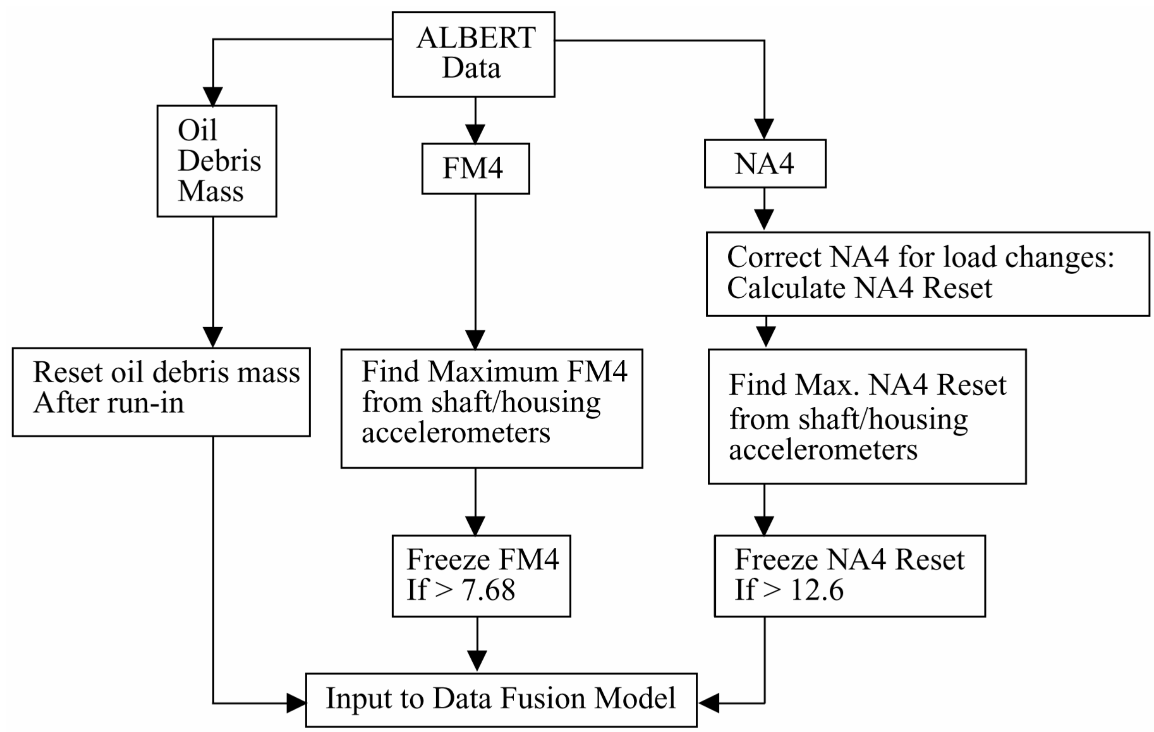


Figure 3.41.—Pre-processing of experimental data prior to input into fuzzy logic model

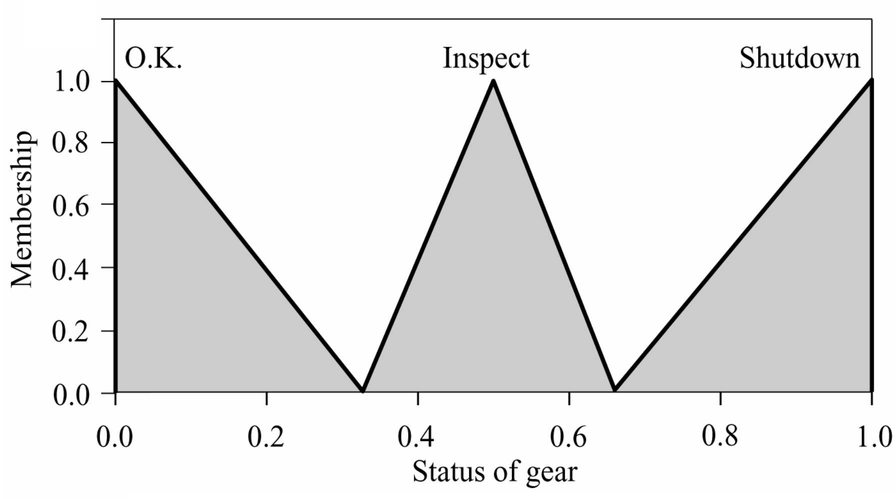


Figure 3.42.—Output of fuzzy logic model

TABLE 3.31
Rules for Fuzzy Logic Model

1. If (FM4 is DL) and (NA4 is DL) and (debris is DL) then (output is O.K)
2. If (FM4 is DH) and (NA4 is DH) and (debris is DH) then (output is SHUTDOWN)
3. If (FM4 is DL) and (NA4 is DL) and (debris is DM) then (output is INSPECT)
4. If (FM4 is DL) and (NA4 is DH) and (debris is DL) then (output is O.K)
5. If (FM4 is DL) and (NA4 is DL) and (debris is DH) then (output is INSPECT)
6. If (FM4 is DH) and (NA4 is DL) and (debris is DL) then (output is O.K)
7. If (FM4 is DH) and (NA4 is DL) and (debris is DM) then (output is INSPECT)
8. If (FM4 is DH) and (NA4 is DH) and (debris is DL) then (output is INSPECT)
9. If (FM4 is DH) and (NA4 is DL) and (debris is DH) then (output is SHUTDOWN)
10. If (FM4 is DH) and (NA4 is DH) and (debris is DM) then (output is INSPECT)
11. If (FM4 is DL) and (NA4 is DH) and (debris is DH) then (output is SHUTDOWN)
12. If (FM4 is DL) and (NA4 is DH) and (debris is DM) then (output is INSPECT)

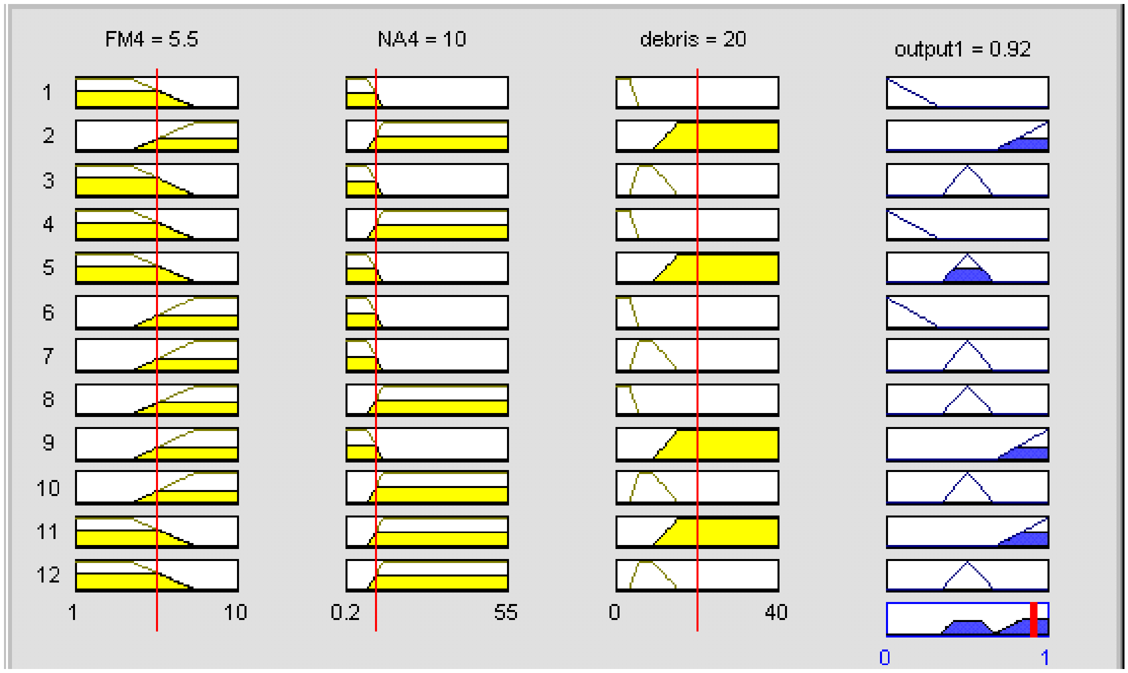


Figure 3.43.—Output of rules for each membership function

Experiment 2 is plotted on Figure 3.44. Destructive pitting was first observed on one tooth of the driven gear at reading 2199 and the output plot indicated to inspect the gears. As the damage increases, the inspect state changes to the shutdown state for this experiment. Experiment 3 is plotted on Figure 3.45. Destructive pitting was first observed on two teeth of both the driven and driver gear at reading 2669 and the output plot indicates to inspect the gears. As the damage increases, the inspect changes to shutdown for this experiment. Experiment 12 is plotted on Figure 3.46. No signs of pitting were observed during this experiment, and the output plot remains in the O.K. region. Experiment 8 is plotted on Figure 3.47. Initial pitting was first observed on 2 driver teeth and 3 driven teeth at reading 5181 and the output plot indicates to inspect the gears. As damage increases, inspect changes to shutdown for this experiment.

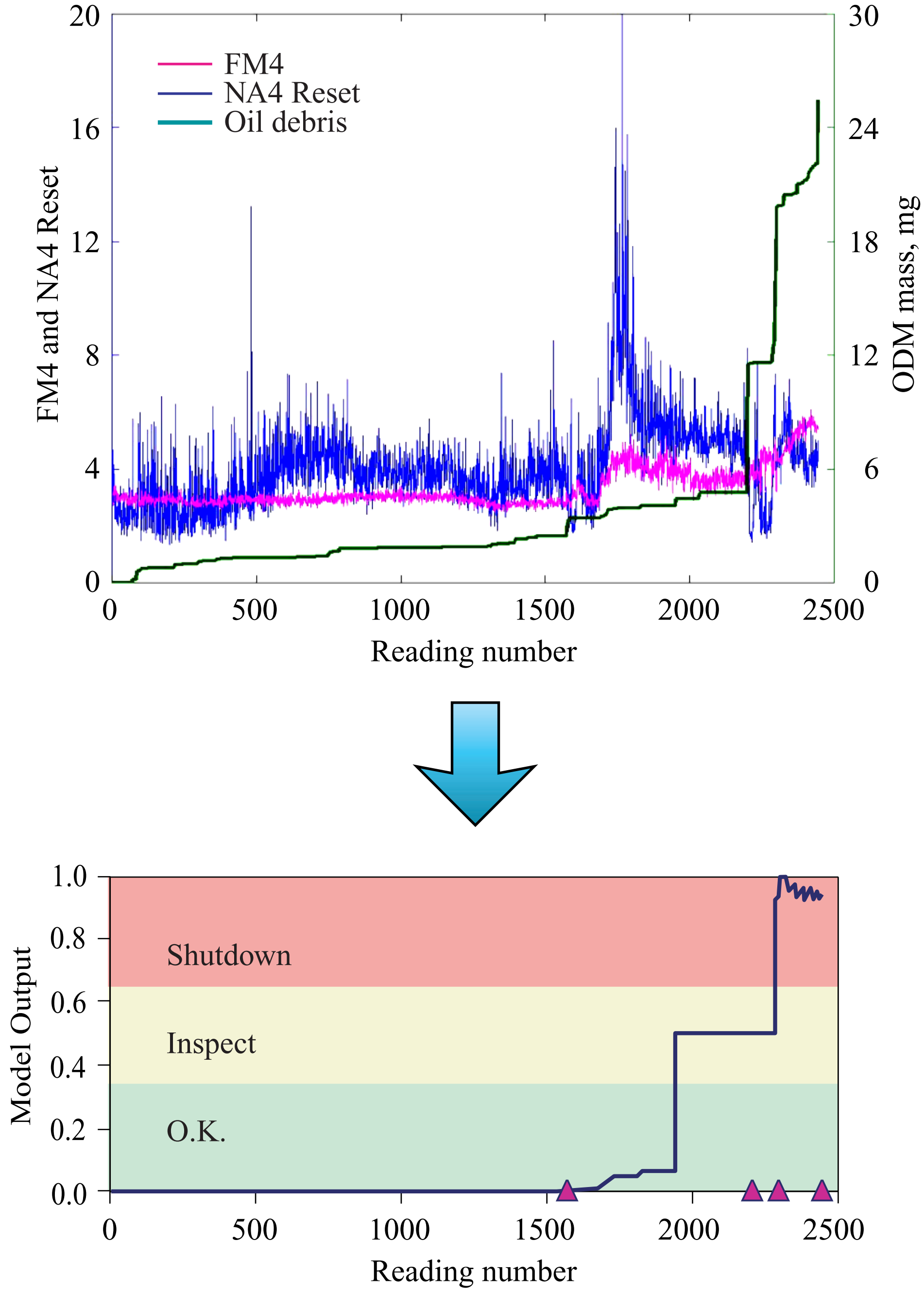


Figure 3.44.—Experiment 2 features and model output

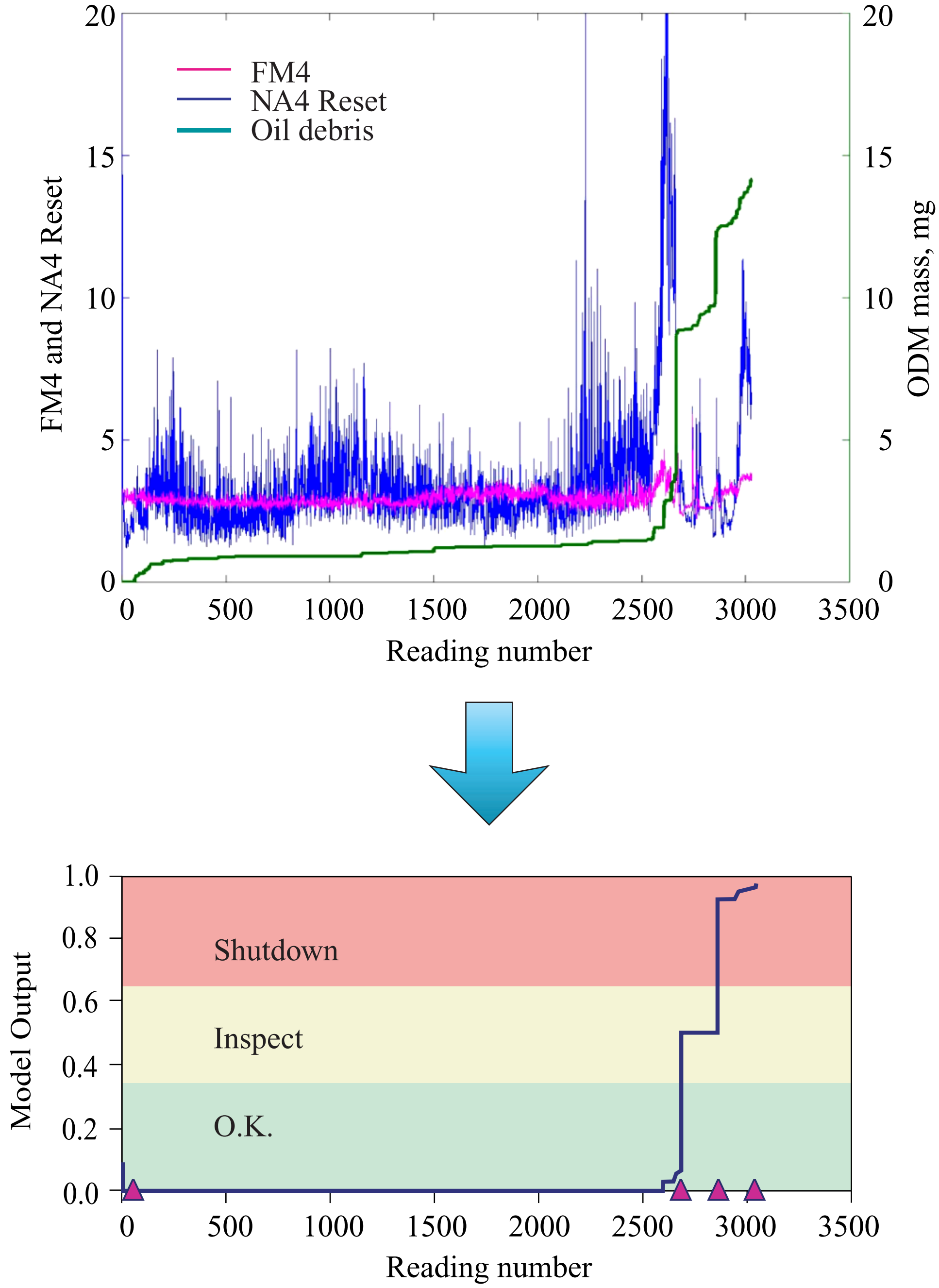


Figure 3.45.—Experiment 3 features and model output

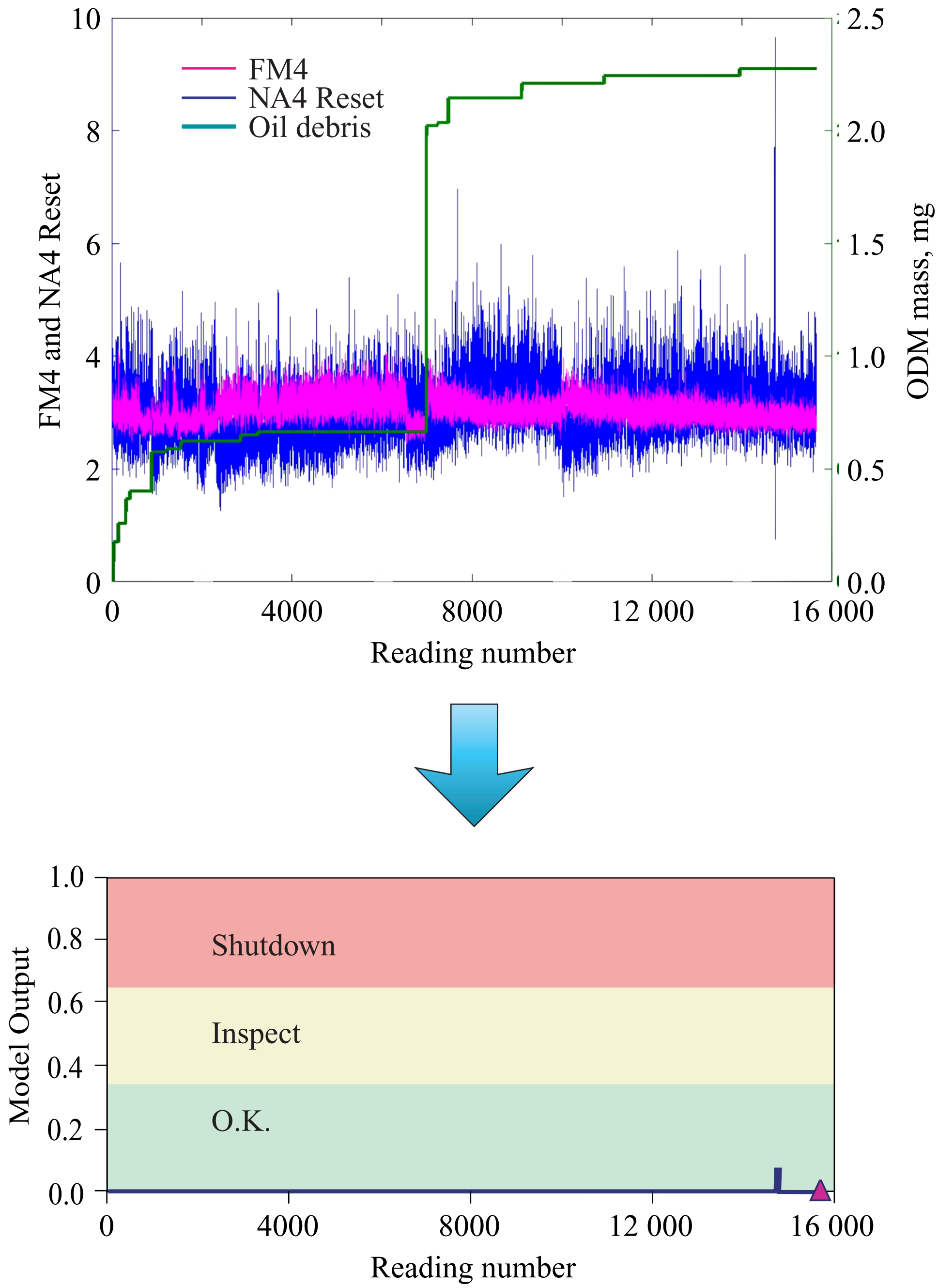


Figure 3.46.—Experiment 12 features and model output

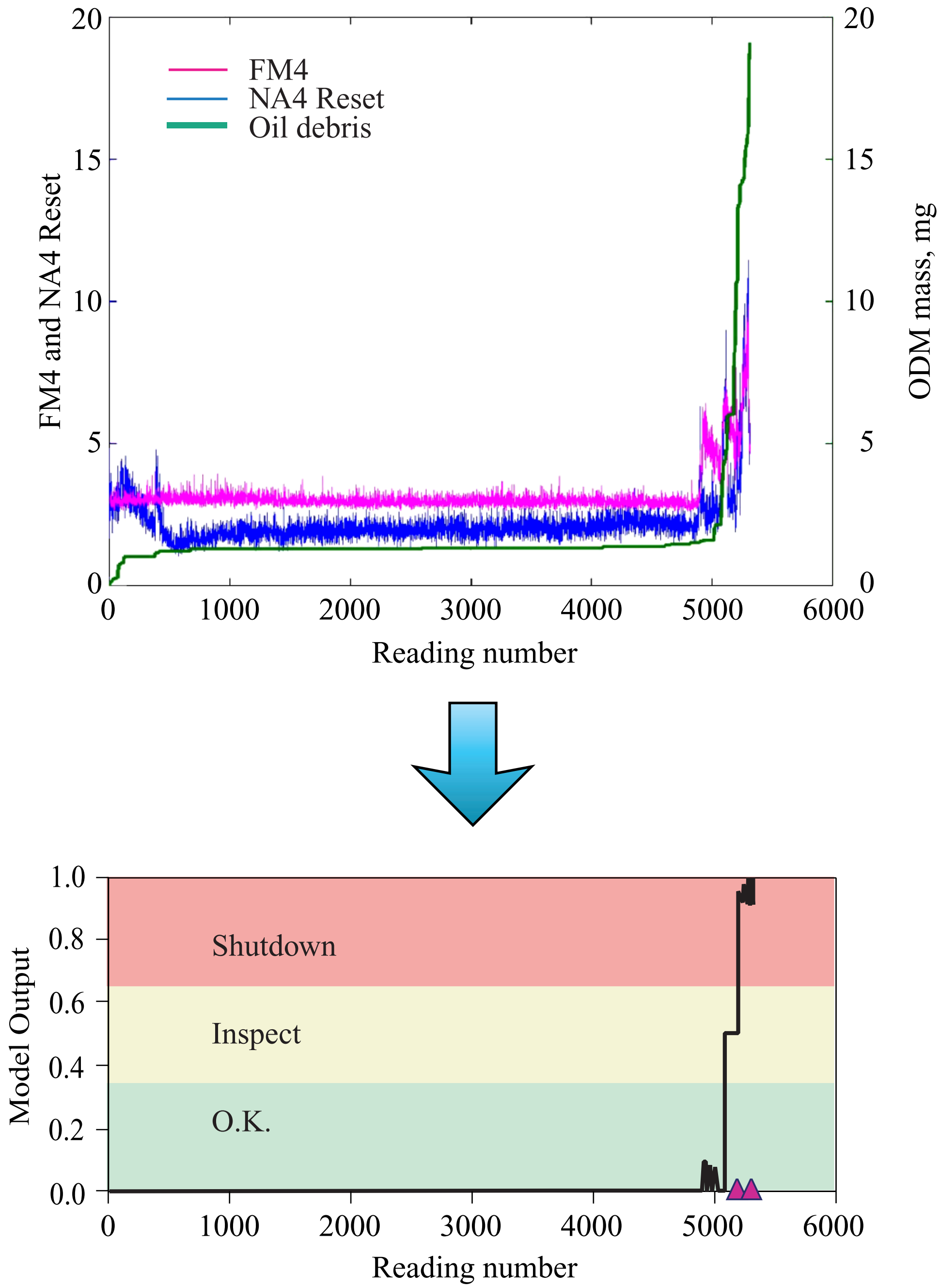


Figure 3.47.—Experiment 8 features and model output

Chapter 4

RESULTS AND DISCUSSION

4.1 Assessment of Diagnostic Features Integration

The advantage of integrating the features of different measurement technologies into a simple fuzzy logic model is evident from a review of the experimental results of this research. The analysis output gives clear information to the end user when making a decision based on the data. The feature integration method developed in this research incorporates the expert knowledge of the diagnostician into a system that can be used to make clear decisions on the status of the geared system.

Several observations are worth noting after careful analysis of the data. The first is that the oil debris feature was more reliable than the vibration features for detecting pitting fatigue failure of spur gears. The vibration features were more sensitive to the environment (operational effects, location, sampling rates, etc.) and these sensitivities are more difficult to quantify or correct for in the field. The vibration algorithms were chosen because the literature had shown they were successful for detecting pitting damage on gears (Zakrajsek (1993); Zakrajsek, et al. (1994a); Zakrajsek, et al. (1994b); and Zakrajsek, et al. (1995b)). Some literature noted NA4 was affected by load change, but the magnitude this effect had on the performance of the algorithm was not discussed (Zakrajsek, et al. (1994a); (Zakrajsek, et al. (1995a); and Zakrajsek (1994)). This operational effect had to be corrected for in this study to maintain the integrity of this algorithm. NA4 Reset was developed during this research as the result of this operational effect. The false alarm rates of NA4 without this correction would be high.

Another observation is that a technique for setting accurate threshold limits for vibration algorithms was not clearly defined in the literature (Zakrajsek (1993); (Zakrajsek (1989); Zakrajsek, et al. (1994a); Zakrajsek, et al. (1995a); Zakrajsek (1994); and Zakrajsek, et al. (1995b)). Setting thresholds appears to be a trial and error method that changes for each experiment and each test rig. This makes it very difficult to quantify the false alarms and missed hits using the individual algorithms. If the threshold limits for the vibration algorithms are set at any number above the nominal value of 3.0, false alarms would dominate (Stewart (1977); Zakrajsek (1993); Zakrajsek (1989); and Zakrajsek (1994)). Why is it so difficult to set limits for these algorithms? Due to limited damage data, developers of vibration diagnostic methods collect data, then when a failure occurs, the diagnostician looks at the data to see if and when the vibration pattern changed. Then a few more data sets are run in a controlled environment looking for this vibration pattern. A specific limit is not defined, rather the diagnostician has to diligently look for a change in the algorithm output that did not previously exist. It is very difficult for an end user, who is unfamiliar with the tests conditions, to differentiate between the change due to damage as compared to the change due to operational effects.

In comparison, the thresholds for this analysis were determined based on membership functions defined for 17 experiments with varied operational conditions. The process used to define membership functions for the vibration algorithms was an attempt to intelligently define threshold limits. Setting thresholds using membership functions is an improvement over the current trial and error method, and gives the end user more flexibility in defining threshold limits based on levels of damage. However, this method also has its limitations in that it requires several sets of damage data to refine the limits.

Like threshold limits and operational constraints, metrics to quantify improvements of new diagnostics tools over existing systems is noticeably absent from the literature. This made it very difficult to quantify the improvement of this system over current HUMS. The only reference found that refers to the false alarm rate of current health monitoring system states false alarms range 1 per 100 flight hours (Stewart (1997)). Of the two references found that refer to the damage detection rates of current health monitoring system, one states a 60 percent damage detection rate, the other claims a 70 percent damage detection rate (Stewart (1997) and Larder (1999)). These references do not discuss data collection methods and rates. If data were collected during steady state conditions, then the number of false alarms would be lower than the false alarm rates under changing operational parameters. Increased frequency of data collection increases the damage detection rates but may increase the false alarm rates.

Because of the limited performance metrics of existing diagnostic tools, assessment of the fused system was not a simple task. A simple, conservative approach was chosen to quantify the diagnostics benefits of the fused system over the individual features. This approach was based on several assumptions:

1. Focus on missed hits and false alarms for destructive pitting failures. A missed hit is when destructive pitting occurs on one or more teeth but was undetected by the feature. A false alarm is when destructive pitting did not occur but is indicated by the feature.
2. For the vibration features, only look at the maximum value between inspection intervals. In other words, if the maximum value exceeds the limit within an inspection period with no damage, only one false alarm occurs per experiment. In reality, the feature can exceed its limit numerous times in the inspection interval causing many false alarms.
3. The false alarms and missed hits will only be determined for each experiment, not within each experiment.

First, the performance of the fused features will be discussed. During experiments 1 to 17 used to create the model, destructive pitting damage occurred during 7 of the experiments. No false alarms or missed hits occurred during these 7 experiments. Experiment 7 was not used in this analysis because only initial pitting occurred for this experiment. One false alarm occurred for the 4 of the experiments with no damage. Only 4 experiments were analyzed because NA4 Reset could not be calculated during experiments 13 to 17 because load was not measured during these tests.

Based on membership function limits of the 16 experiments, a comparison will be made between the performance of the fused features, the two vibration features, and the oil debris feature. Refer to Tables 3.5 to 3.7, Tables 3.25 to 3.30, and Figures 3.23, 3.39, and 3.40 for the oil debris mass, FM4, and NA4 Reset maximum values during inspection intervals. Table 4.1 shows the results of assessing false alarm rates and missed hits for the

TABLE 4.1
Missed Hits and False Alarms

Feature	Limits	Experiments with Damage		#	Experiments with No Damage	#	Missed Hits	False Alarms
		Missed Hits	False Alarms		False Alarms		%	%
FM4	4.04	2	0	7	7	9	29	44
FM4	7.36	6	0	7	2	9	86	13
NA4 Reset	7.86	2	2	7	3	4	29	45
NA4 Reset	12.6	4	1	7	1	4	57	18
Oil Debris Mass	5.45	0	4	7	1	9	0	31
Oil Debris Mass	8.69	0	3	7	0	9	0	19
Oil Debris Mass	15.47	5	0	7	0	9	71	0
Fused Features		0	0	7	1	4	0	9

data used to build the model. The first column defines the feature. The second column is the limit. If the feature value is greater than or equal to the limit, damage is indicated. The third and fourth columns show the number of missed hits and false alarms for the experiments with damage. The fifth column is the number of experiments with damage. The sixth shows the number of false alarms for experiments with no damage. The seventh column shows the number of experiments with no damage. The "missed hits" column lists the percentage of missed hits for the 7 experiments with damage. The "false alarm" column is the percentage of false alarms based on 16 experiments for FM4 and oil debris mass, and 11 experiments for NA4 Reset. From this table, the dilemma faced by the facility operator, maintenance person, or pilot is obvious. There is a tradeoff between the sensitivity of the system to detect damage and the number of false alarms. If the limit is decreased, less missed hits and more false alarms will result and if the limit is increased more missed hits and less false alarms will occur.

How does the performance of the individual features compare to the fused output results for these experiments? For the 7 experiments with pitting damage no false alarms or missed hits occurred when using the fused output. For the 4 experiments with no damage (experiments 9 to 12) 1 false alarm occurred for experiment 10. The fused features have the lowest false alarm rate. The results of this research show the benefit of combining two measurement technologies and 3 features using the data set to develop the data fusion model. But, the important question is what happens when new data, not used to develop the model is used? Can the model continue to detect damage? The membership functions were developed based on data collected during experiments 1 to 17. Experiments 18 to 24 were new data not used for model development. The data fusion results of Experiments 18 and 24 are shown in Figures 4.1 and 4.2. The features in the plot on the top are displayed on-line during the experiment. The data fusion analysis is done during post-processing. The data from these 2 experiments and the other

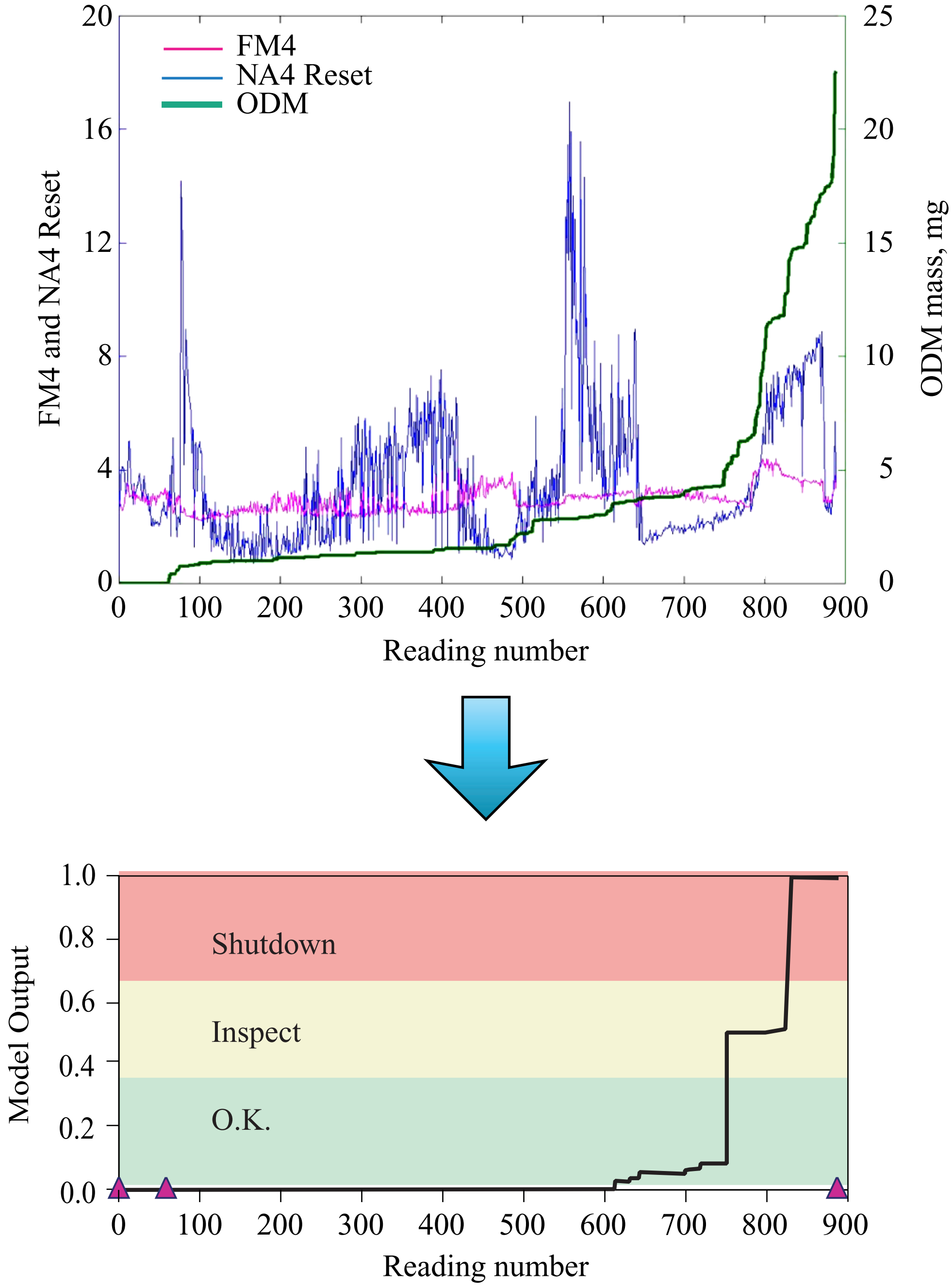


Figure 4.1.—Experiment 18 features and model output

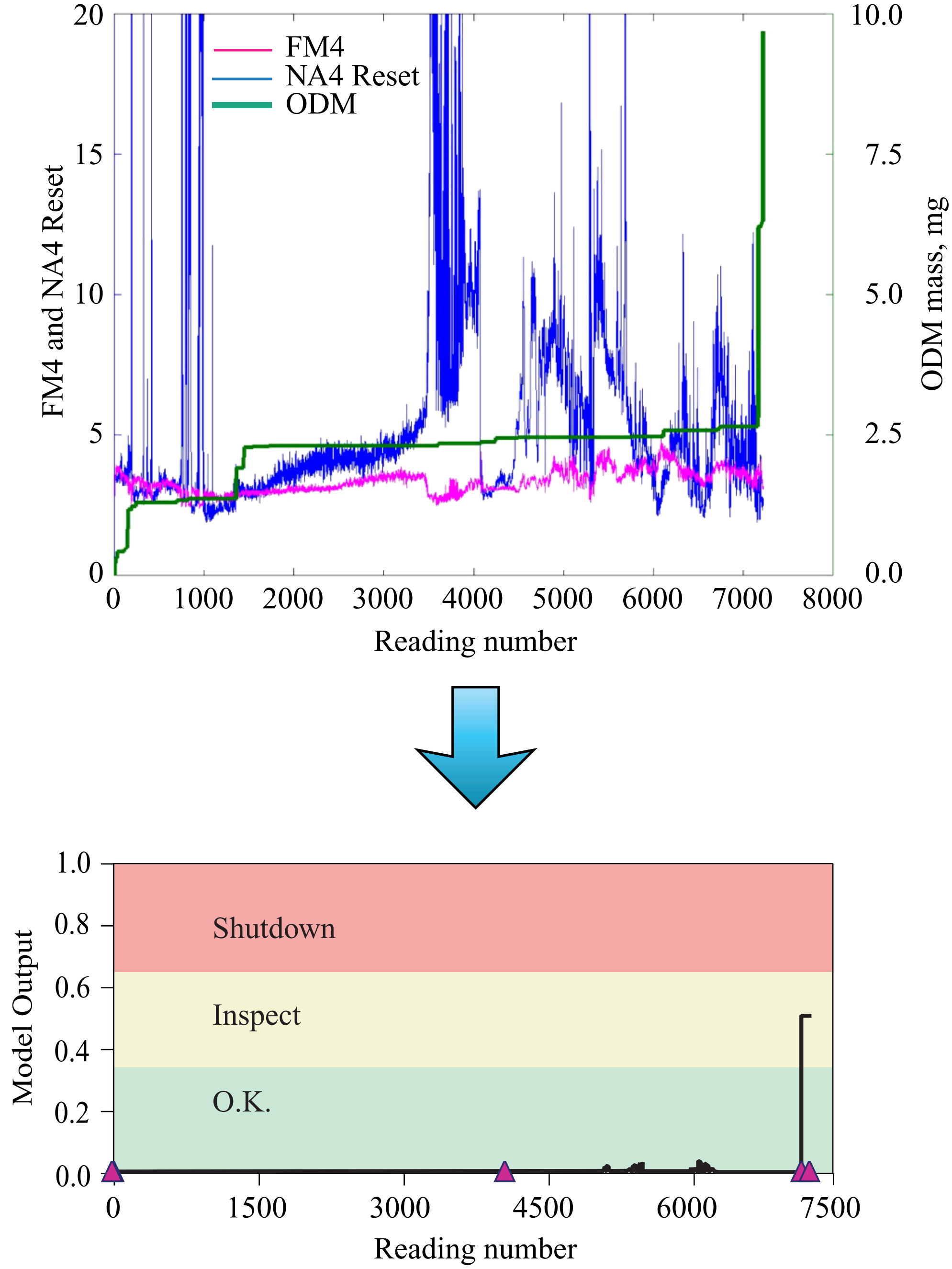


Figure 4.2.—Experiment 24 features and model output

7 experiments with damage indicate the rig should be shutdown for inspection or damage. No false alarms or missed hits were indicated for these 7 experiments based on the fused output.

This research, focusing on pitting damage, found the oil debris feature more reliable than vibration features. But, during one experiment, a tooth cracked on the driven gear without warning. Oil debris analysis is unable to detect cracks or fractures that do not release significant debris (Astridge (1987)). At test completion, minimal initial pitting was barely visible with the naked eye on the other teeth on both the driver and driven gear. The oil debris in the rig increased at the beginning of the test due to an operational change, but did not change significantly until after the tooth cracked. Results from crack propagation tests in the NASA Glenn Spur Gear Fatigue Rig indicated NA4 reacted to fatigue cracks in standard spur gears (Zakrajsek, et al. (1996)). For this reason, NA4 Reset data was analyzed to determine if it detected the cracked tooth.

Figure 4.3 is a plot of the vibration and oil debris features with an expanded scale near the time the tooth cracked. Referring to Figure 4.3 for analysis of this experiment, video inspection was performed at reading 6393 with no pitting visible. At shutdown, reading 7148, tooth 18 of the driven gear was cracked. Figure 4.3 shows the vibration feature from the video inspection reading until shutdown, and an expanded scale showing the two NA4 Reset spikes prior to the debris increase, suggesting that NA4 indicated a cracked tooth prior to the debris increasing. The time-synchronous average data were also plotted to determine whether the increase in NA4 Reset was attributed to an increase in magnitude at this tooth location. A cracked or broken tooth is often detected in the time

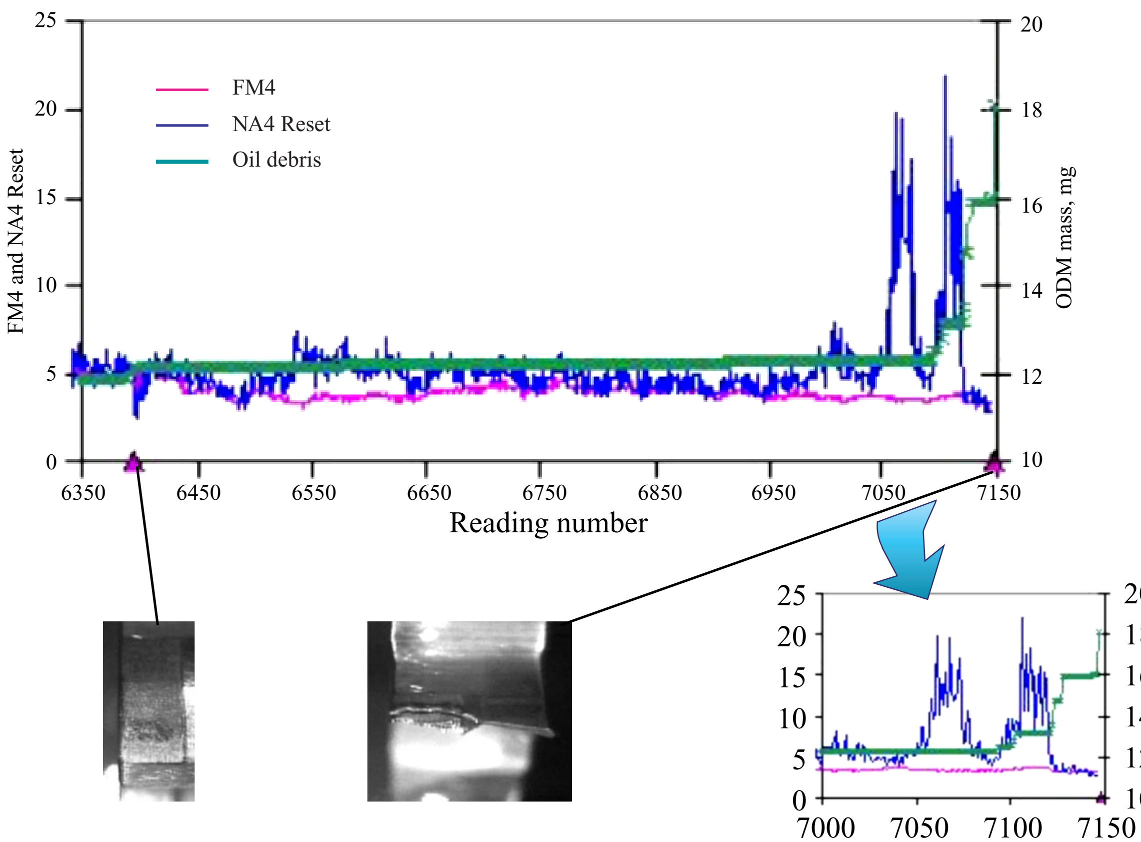


Figure 4.3.—Experiment with cracked tooth

waveform which shows a spike at the once per revolution of the cracked tooth (McFadden (1987)). The amplitude at tooth 18 showed a significant increase at approximately reading 7061 as compared to previous readings. This would indicate NA4 detected the cracked tooth approximately an hour before the oil debris sensor. Since it was based on only one data set, this does not prove NA4 Reset is the best method for crack detection. It only reinforces the importance of using several different measurement technologies for damage detection, to recognize unanticipated failure modes, since all have strengths and weaknesses for different applications.

4.2 Application of Data Fusion Method to Other Systems

The work conducted in this study thoroughly characterized the effect spur gear pitting fatigue has on vibration algorithms FM4, NA4 Reset and oil debris. The diagnostics developed have improved the performance of fatigue tests conducted in the NASA Glenn Spur Gear Fatigue Test Rig. Additional data is required to implement this system on other systems. Although vibration data has been collected on other systems, oil debris mass failure data has been limited to the spur rigs. The oil debris sensor manual and other research on ball bearings indicate the debris mass detected is proportional to the rolling element diameter and the pitch diameter (Metalscan User's Manual; Miller and Kitaljevich (2000)). Does a gear relationship exist that can be used to modify the oil debris membership functions limits when failure data are acquired on other systems?

Due to the success of oil debris analysis in predicting damage on the Spur Gear Fatigue Rigs, an oil debris sensor was installed on the NASA Glenn Spiral Bevel Gear Test Facility. Spiral bevel gears are used in helicopter transmissions to transfer power between nonparallel intersecting shafts. A detailed description of this test facility can be found in Handschuh (1995, 2001). The Spiral Bevel Gear Test Facility is illustrated in Figure 4.4.

The main purpose of this test rig is to study the effects of gear material, gear tooth design, and lubrication on the fatigue strength of gears. The facility uses a closed loop torque regenerative system. Two sets of spiral bevel gears are tested simultaneously. Fatigue tests are performed on aerospace quality gears under varying operating conditions. The 12 tooth pinion and 36 tooth gear have 5.14 in. (13.06 cm) diametral pitch, 35 degree spiral angle, 1 in. (2.54 cm) face width, 90 degree shaft angle, and 22.5 degree pressure angle. Tests are performed for a specified number of hours or until surface fatigue occurs.

Four experiments were performed. Oil debris data were collected during all four experiments from the same type of oil debris sensor used in the spur rig, but a larger size (3/4"). The larger size was required due to the higher oil flow rates in this rig. The 3/4" sensor is less sensitive than the 3/8" used in the spur rig. This sensor required the bins to be reconfigured from 16 to 14 bins with the smallest particle detected at 225 microns. The bin sizes are listed in Table 4.2.

The instrumentation available varied for each of the four experiments due to test schedule constraints. For the first and second experiments, an oil debris sensor was installed on the rig, but accelerometers were not installed. For the second experiment, damage progression data during the experiment was not collected due to the unavailability of the data acquisition system. The data for the second experiment was

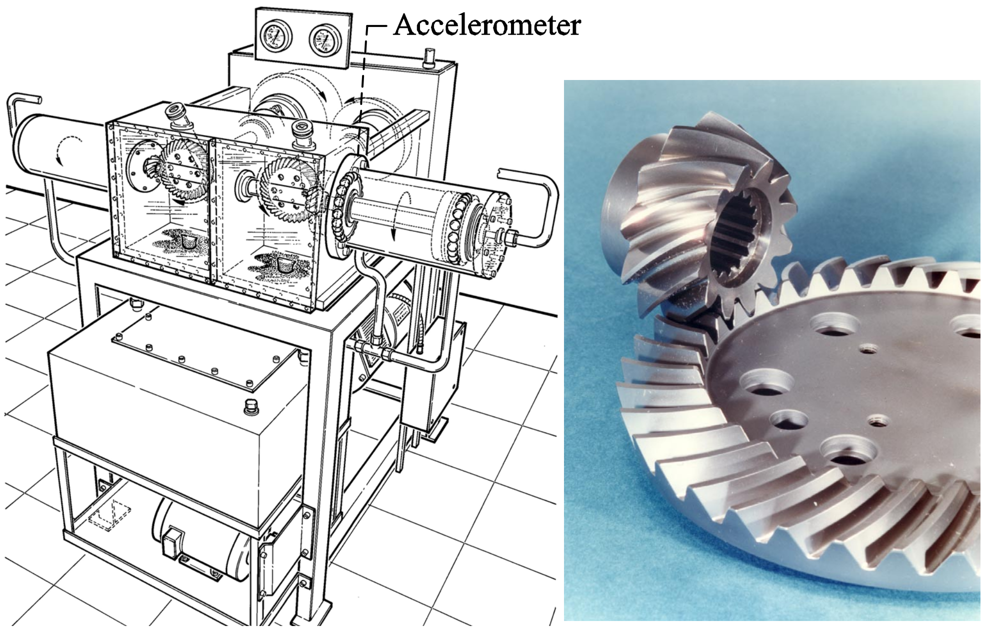


Figure 4.4.—Spiral bevel gear fatigue rig.

TABLE 4.2
Oil debris particle size ranges for bevel tests

Bin	Bin range, μm	Average	Bin	Bin range, μm	Average
1	225–275	250	8	575–625	600
2	275–325	300	9	625–675	650
3	325–375	350	10	675–725	700
4	375–425	400	11	725–775	750
5	425–475	450	12	775–825	800
6	475–525	500	13	825–900	862.5
7	525–575	550	14	900–1016	958

limited to the total accumulated oil debris mass at test completion. For the third and fourth experiments, accelerometers were installed on the right and left pinion shaft bearing housings to measure vibration. The location of the right accelerometer is shown in Figure 4.4. The left accelerometer was placed at the same position on the left side of the gearbox. The accelerometers are lightweight, piezoelectric accelerometers, of the same type used on the spur gear shaft. Shaft speed was also measured with the same type of optical sensor used in for the spur gear tests.

The test pinion had 12 teeth with a shaft speed of 10,200 RPM and the gear had 36 teeth with a shaft speed of 3400 RPM. The meshing frequency was 2040 cycles/second. The shaft speed was measured on the test gear shaft. For every revolution of the test gear, there were three revolutions of the pinion. Data was sampled for 100KHz for 2 seconds duration. Time-synchronous averaging were performed for 113 revolutions of the test gear.

The oil debris mass generated during the first bevel rig experiment with pitting damage is shown on Figure 4.5. At test completion, initial pitting was observed on one pinion tooth. An image of this damage is also shown on Figure 4.5. Can this amount of debris be somehow correlated to the debris measured during spur rig test? Does a simple correlation exist between gear tooth contact area and amount of debris? If so, it would be very easy to change the membership functions of the oil debris feature proportionally for different systems. As discussed in section 3.1.4, Oil Debris Feature, 25% of the tooth surface contact area for one spur gear tooth was calculated to get a feel for the amount of damage related to gear tooth size. If .0397 cm diameter pits densely covered the surface contact area of one spiral bevel gear tooth, it results in 53.6 mg, which is 5.95 times the debris calculated for the spur gears. This factor was multiplied by the limits defined in the oil debris membership functions. Figure 4.6 shows the new membership functions for levels of damage for the bevel gears. Table 4.3 shows the oil debris feature values for the spur and bevel rigs. The oil debris mass is input into the oil debris membership function and the resulting output is shown in Figure 4.7. Applying this simple change to the membership functions, the model indicates inspection should be performed on the bevel gears. Since initial pitting was first observed at reading 1028 with a mass of 8.771 mg measured, these values may have to be adjusted as additional data are collected. However, as a rule of thumb, this is a good value to start with when setting oil debris damage levels.

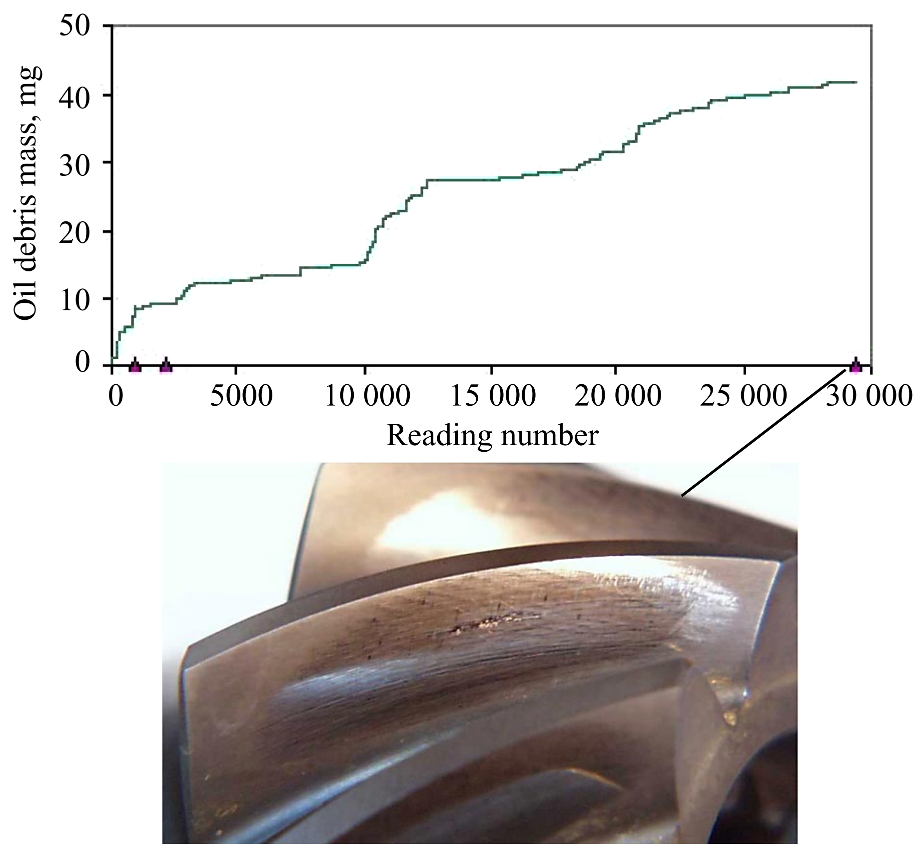


Figure 4.5.—Oil debris mass measured during bevel test 1.

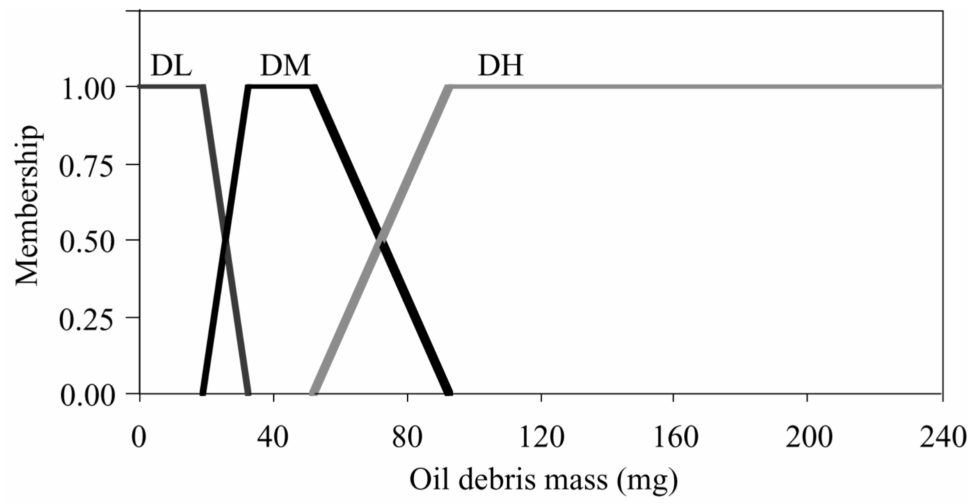


Figure 4.6.—Bevel rig membership functions for levels of damage.

TABLE 4.3
Spur rig and bevel rig membership function values

DL Damage Low			DM Damage Medium			DH Damage High		
Spur	Bevel		Spur	Bevel		Spur	Bevel	
0	0	1	3.159	18.8	0	8.69	51.7	0
3.159	18.8	1	5.453	32.4	1	15.475	92.1	1
5.453	32.4	0	8.69	51.7	1	40	238	1
			15.475	92.1	0			

During the second experiment on the bevel rig, the data acquisition system was unavailable, but the oil debris sensor monitored the debris that accumulated throughout the test. The experiment lasted 565 minutes. At test completion, 104.3 mg of debris was measured. This includes the run-in debris generated during the test. Damage occurred on both the right and the left pinion. Figure 4.8 shows the amount of damage to the right pinion gear teeth. Destructive pitting occurred on two teeth. Figure 4.9 shows the amount of damage to the left pinion gear teeth. Wear marks and the start of initial pitting also occurred on one tooth. Looking at the oil debris membership function values shown in Figure 4.5, 104.3 mg falls within the Damage High region of the oil debris membership function.

Vibration data and oil debris data were collected for the third bevel experiment. FM4 and NA4 Reset were calculated from the vibration data for this experiment. Since FM4 was based on normalized statistical functions, the FM4 membership functions developed for the spur gear tests could be applied to this system. It was not clear if NA4 Reset membership functions could be used. NA4 Reset is calculated by correcting for load fluctuations. When a 10% load fluctuation occurred during spur rig tests, the denominator for NA4 was reset. The vibration data collected during the bevel gear tests were even more sensitive to load fluctuations. NA4 Reset was calculated for the bevel tests by correcting for a 4% load fluctuation. This enabled the same NA4 Reset membership functions used for the spur gear tests to be used for the bevel rig tests. The membership values for the FM4 and NA4 Reset features were shown on Figures 3.39 and 3.40.

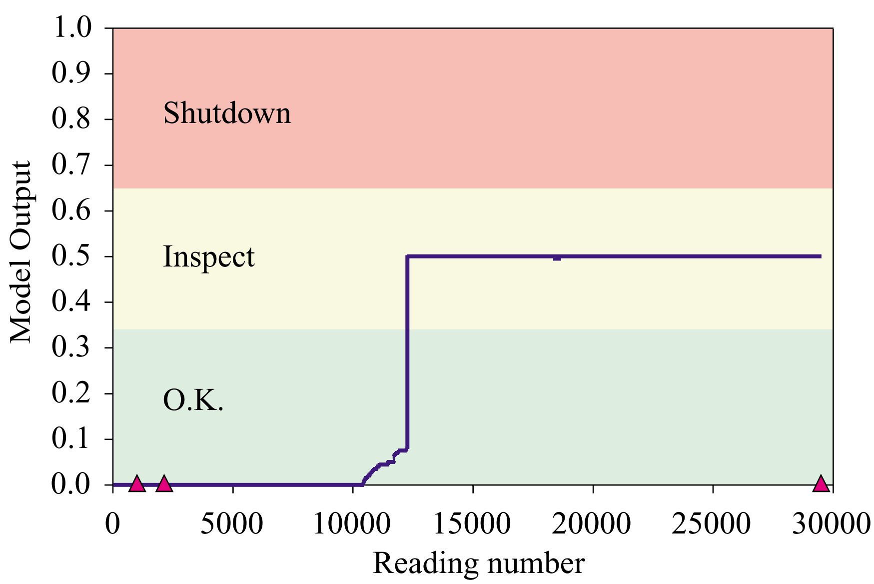


Figure 4.7.—Output of oil debris membership function for bevel test 1.

Figure 4.8.—Damage to right pinion teeth during bevel test 2.

Figure 4.9.—Damage to left pinion teeth during bevel test 2.

Vibration data were collected on both sides of the gearbox. However, the vibration feature calculated on the right side began to substantially increase, indicating damage was beginning to occur on the right side of the test rig. The right side vibration and oil debris features are plotted on Figure 4.10. The gears were inspected at test completion and damage to two teeth on the right side pinion is shown in Figure 4.11. Figure 4.12 shows the output of the fuzzy logic model for this experiment. The only change to the model from the spur gear tests is the change to the oil debris membership values for the 3 levels of damage as shown in Figure 4.6. As discussed in section 3.3, Feature Validation, the maximum value for NA4 Reset and FM4 is input into the model. For the Spur Gear Fatigue Test Rig, the two accelerometers are measuring one pair of meshing gears. For the Spiral Bevel Fatigue Rig, the left accelerometer measures the vibration from the left pair of gears and the right accelerometer measures the vibration from the right pair of meshing gears (See Figure 4.4). Although it is outside the scope of this thesis, a more complex model can be developed that indicates the location of the damage based on the accelerometer location. The data obtained from the Spiral Bevel Fatigue Rig also reinforces the importance and benefit of fusing different measurement technologies for a more reliable damage detection system.

Vibration data and oil debris data were also collected for the fourth bevel experiment. Vibration data were collected on both sides of the gearbox. For this experiment, the vibration feature calculated from the left accelerometer began to substantially increase indicating damage was beginning to occur to the gears on the left side of the test rig. The left side vibration feature and the oil debris feature are plotted on Figure 4.13 and with an expanded scale in Figure 4.14. The gears were inspected at test completion and damage to one tooth on the left side pinion is shown in Figure 4.15. Figure 4.16 shows the output of the fuzzy logic model for this experiment. The initial output indicates inspection at approximately reading 6000. But, looking at NA4 Reset, this feature began to increase after reading 5200. NA4 reset appears a more reliable feature for the bevel rigs as compared to the spur rigs. In order to capture this improved feature performance, one rule in the model was modified. For the Spur Gear Fatigue Rig if the NA4 Reset feature was high, but the FM4 and oil debris features were low, the state of the system was O.K. This output state was changed to Inspect for the bevel tests. The changed rule, rule 4, is highlighted in Table 4.4. The output of the model, with rule 4 changed, is shown in Figure 4.17. This new output indicates to inspect the gears at reading 5400, 10 hours sooner.

This was a simple system analysis on a limited data from the bevel test rigs. Additional data sets are required to modify the model for optimum performance for this new system. This chapter was included to show how simple changes can be made to the basic model for applications to other geared systems. These promising results show how this process may be applied to other systems. As additional data is acquired on other geared systems, more techniques can be developed for applying this process to more complex systems.

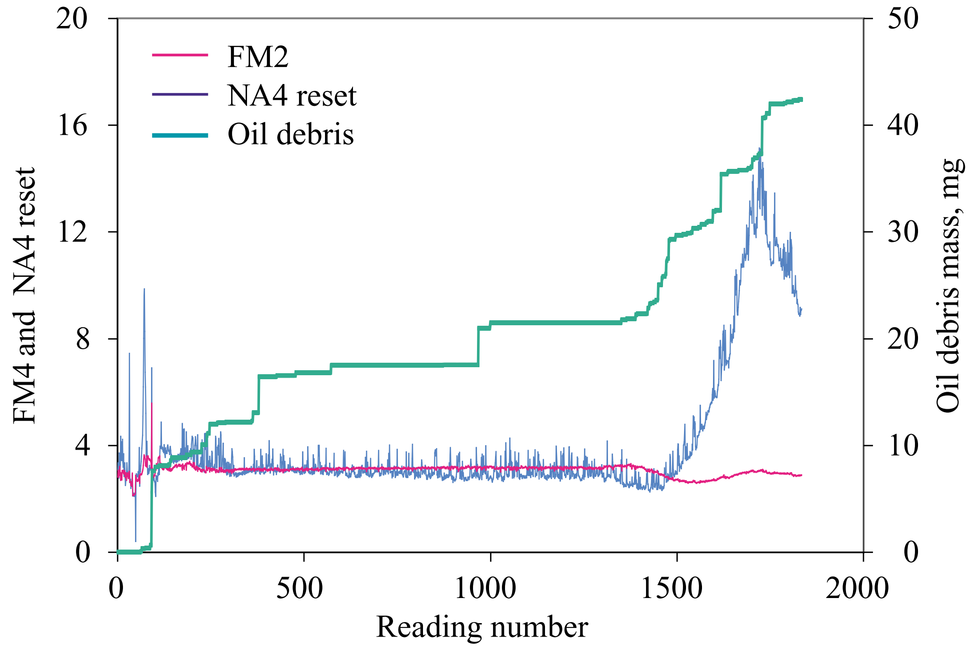


Figure 4.10.—Vibration and oil debris features for bevel test 3.

Figure 4.11.—Damage to right pinion teeth during bevel test 3.

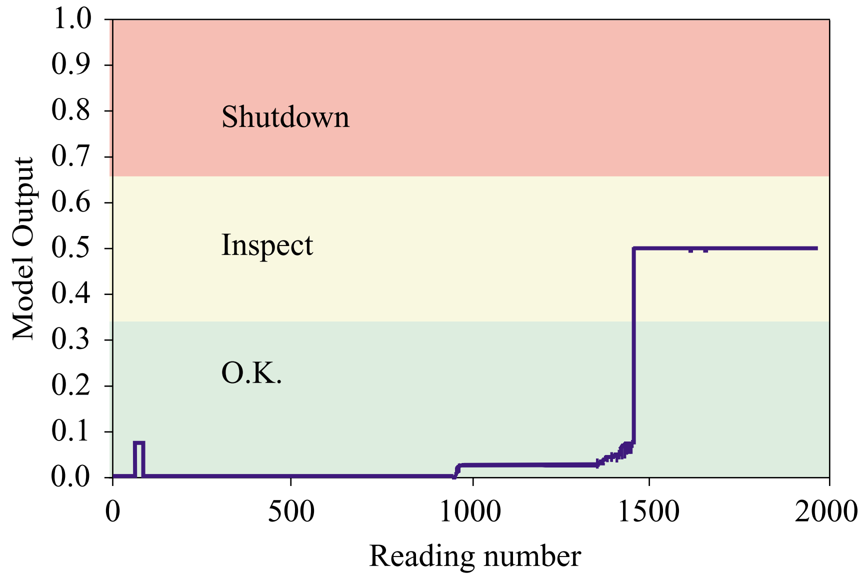


Figure 4.12.—Fuzzy logic model output for bevel test 3.

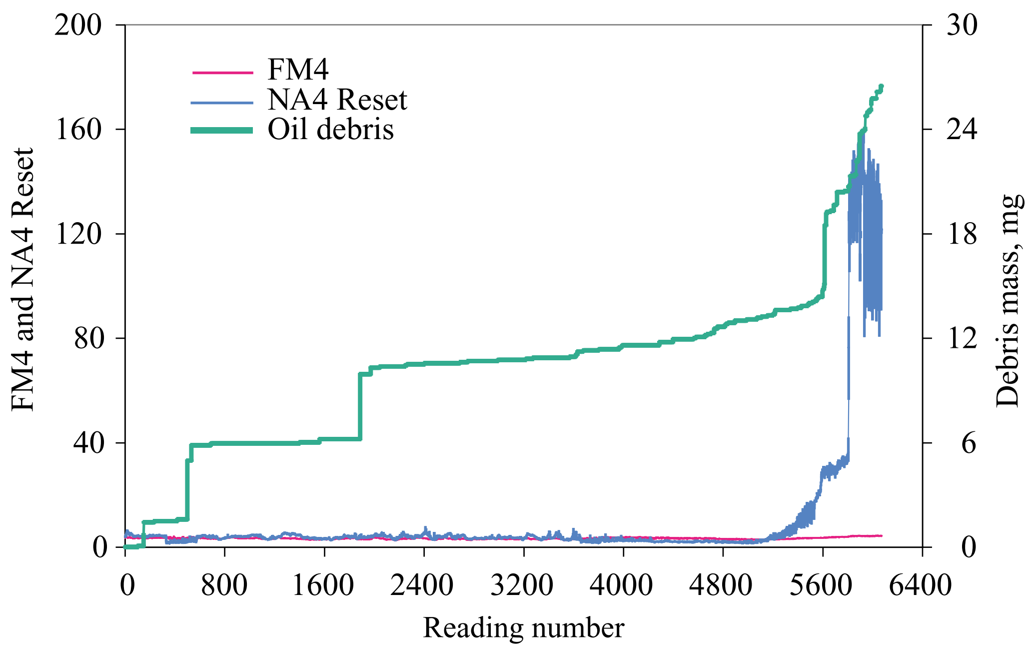


Figure 4.13.—Vibration and oil debris features for bevel test 4.

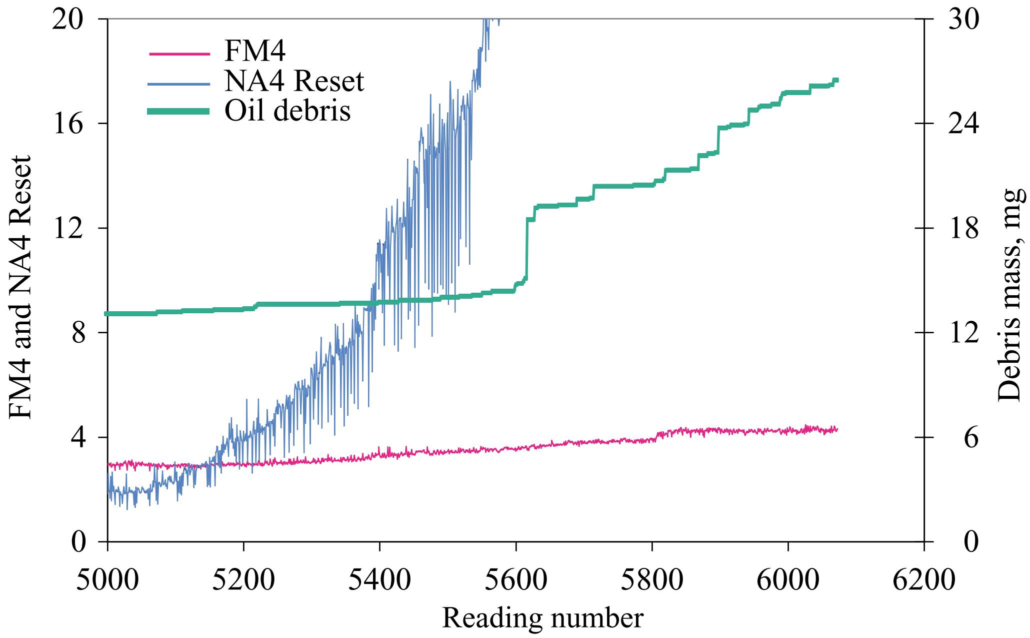


Figure 4.14.—Vibration and oil debris features for bevel test 4 with expanded scale.

Figure 4.15.—Damage to left pinion teeth during bevel test 4.

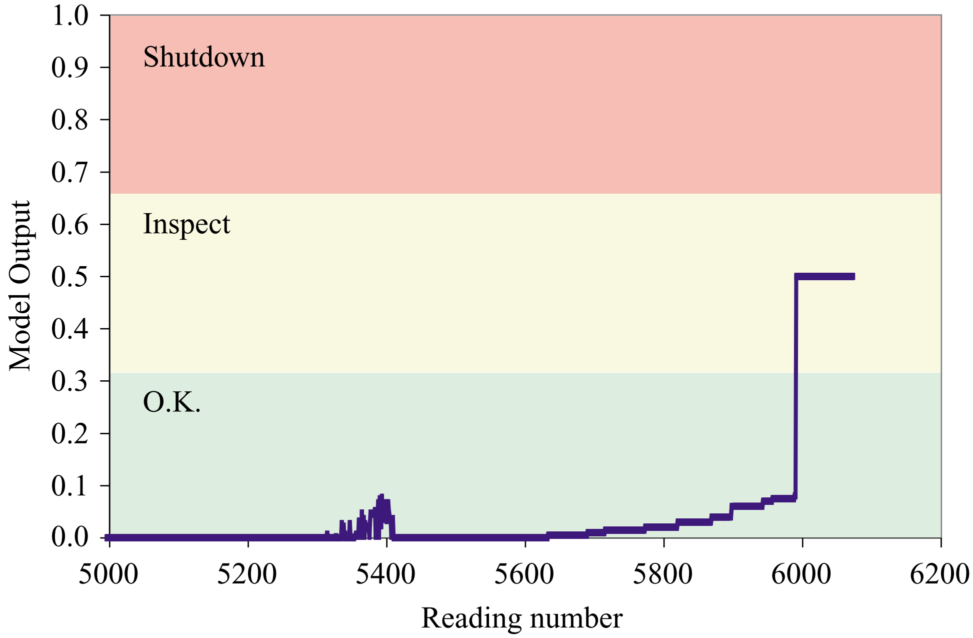


Figure 4.16.—Fuzzy logic model output for bevel test 4.

TABLE 4.4
Rules for Bevel Rig Fuzzy Logic Model

1. If (FM4 is DL) and (NA4 is DL) and (debris is DL) then (output is O.K)
2. If (FM4 is DH) and (NA4 is DH) and (debris is DH) then (output is SHUTDOWN)
3. If (FM4 is DL) and (NA4 is DL) and (debris is DM) then (output is INSPECT)
4. If (FM4 is DL) and (NA4 is DH) and (debris is DL) then (output is **INSPECT**)
5. If (FM4 is DL) and (NA4 is DL) and (debris is DH) then (output is INSPECT)
6. If (FM4 is DH) and (NA4 is DL) and (debris is DL) then (output is O.K)
7. If (FM4 is DH) and (NA4 is DL) and (debris is DM) then (output is INSPECT)
8. If (FM4 is DH) and (NA4 is DH) and (debris is DL) then (output is INSPECT)
9. If (FM4 is DH) and (NA4 is DL) and (debris is DH) then (output is SHUTDOWN)
10. If (FM4 is DH) and (NA4 is DH) and (debris is DM) then (output is INSPECT)
11. If (FM4 is DL) and (NA4 is DH) and (debris is DH) then (output is SHUTDOWN)
12. If (FM4 is DL) and (NA4 is DH) and (debris is DM) then (output is INSPECT)

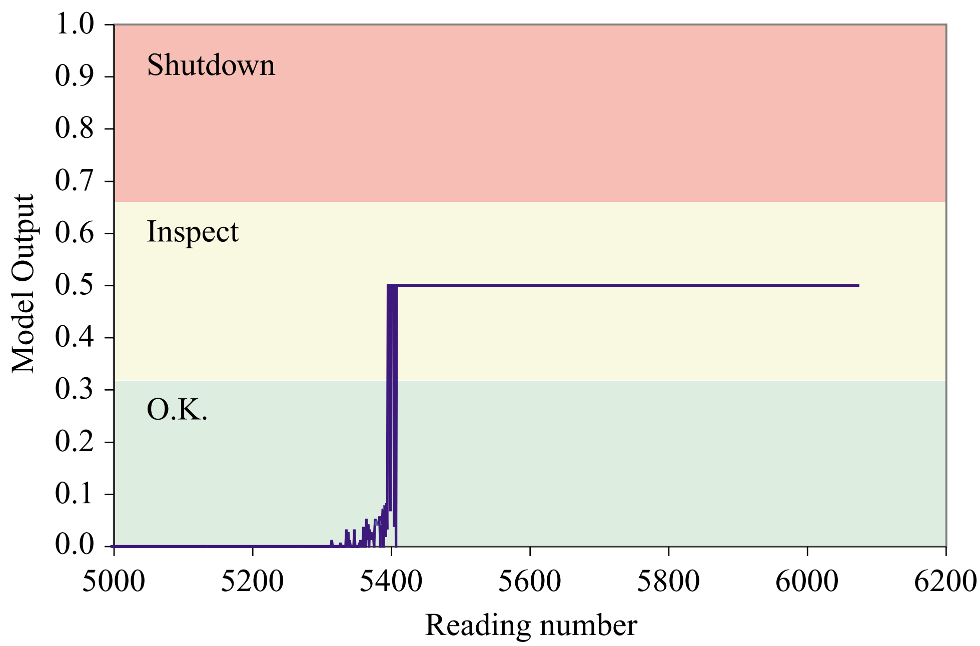


Figure 4.17.—Fuzzy logic model output for bevel test 4 with modified rule 4.

Chapter 5

CONCLUSIONS

The integration of two measurement technologies, oil debris analysis and vibration, results in a system with improved damage detection and decision-making capabilities. Vibration and oil debris data were collected from experiments in the NASA Glenn Spur Gear Fatigue Rig. Oil debris and vibration features were obtained and input into a data fusion process. Using fuzzy logic techniques applied to the oil debris and vibration data, a simple system model was developed that discriminates between stages of pitting wear. Additional tests were run to verify the system detects damage on data not used to build the model. Results indicate combining the vibration and oil debris measurement technologies improves the detection of pitting damage on spur gears. As a result of this research, the diagnostic tools used for damage detection in the NASA Glenn Spur Gear Fatigue Rigs have been significantly improved.

Several other findings were made that will impact the development of health monitoring tools for geared systems. The first being, oil debris analysis is more reliable than vibration analysis for detecting pitting fatigue failure of spur gears. The second finding is that some vibration algorithms are as sensitive to operational effects as they are to damage. NA4 Reset was developed as the result of this finding. The third finding is that vibration algorithms FM4 and NA4 Reset do not indicate damage progression, but the increase in oil debris mass is related to damage progression. The fourth finding is that clear threshold limits must be established by the developer of the diagnostic tool if it is to be applied to other systems. The development of membership functions for each parameter will improve this process. It also enables the end user to replace these parameters with their own by adjusting the membership functions.

The study also identified some of the challenges in the area of diagnostics. First, one of the reasons it is difficult to develop reliable diagnostic tools is the lack of damage data. In many instances, diagnostics are developed based on one or two data sets in controlled environments that do not perform well in the harsh environment a helicopter transmission experiences. A database of continuous transitional data from a normal state to a failed state must be understood. Unfortunately, this data is not readily available and diagnosticians must settle for failure progression data under documented conditions (Becker, et al. (1998)). The data collected in support of this thesis in the NASA Glenn Spur Gear Fatigue Rig will improve the diagnostics used when performing spur gear fatigue tests. Validation in the field under varying environmental conditions is required before benefits to helicopter safety can be claimed. However, the concepts described herein may be applicable to other rotating equipment using other measurement technologies.

The second point is also related to the limited damage data. A set of standard metrics to quantify the performance of each diagnostic tool does not currently exist.

Attempts are being made to develop a comprehensive system for evaluating the performance of vibration based diagnostic tools. A web based prototype application is under development using different metrics to assess vibration algorithms (Orsagh, et al. (1999)). A database of diagnostic data on different systems under different operating conditions is still needed to use this application for metric assessment. Once this work is complete, a fused system can be developed using many different algorithms based on their individual strengths and weaknesses.

The third point relates to the human factors aspect of diagnostic tool development. Many research papers have been written on the development of vibration algorithms and the analysis of oil to detect damage to rotating equipment. As a diagnostician, it is important to identify the end user of the diagnostic tool early on in the process. The end user of most helicopter transmission diagnostic tools is the technician that determines if maintenance is required on the transmission. If the tool being developed requires hours of analysis and large amounts of stored data to determine the health of the system, it is probably not feasible for this application. It is thus important to keep the man-machine interface in mind when designing diagnostic tools.

Future work is planned to implement the health monitoring techniques developed to other drives systems test rigs at NASA Glenn Research Center. Preliminary tests on the Spiral Bevel Gear Fatigue Rig suggest that this work can be successfully implemented on other systems. Once oil debris and vibration failure data is obtained from different geared systems, a parametric analysis can be performed that will enable the development of an optimized system for new applications including different types of gears and different test rigs. Analyzing different measurement technologies for integration is also needed that may enable different types of failure mechanisms to be detected. In a more complicated system with many sensors the membership functions could also be expanded to include sensor failure as a state of the system. This will aid in the troubleshooting of more complex systems.

Probabilistic methods were not utilized in this study primarily because of the unavailability of sufficient damage data. As the system under investigation becomes more complex, and additional damage data are acquired on these complex systems, probabilistic methods will prove beneficial. Probabilistic methods will provide the tools required to classify different types of faults detected on different systems. The future use of probabilistic methods will provide a more overall decision support system in the development of future diagnostic systems.

Results of this research leads to several significant conclusions that will impact the design of future transmission health monitoring systems. The first being measurement of the accumulated mass of the debris generated in the oil is an effective method to predict gear pitting damage. Secondly, data fusion utilizing fuzzy logic analysis techniques can be successfully used to establish alarm limits on the state of the geared system with a decrease in false alarms over conventional trial and error methods. Thirdly, by fusing different measurement technologies and including expert knowledge of the diagnostician into the system, clear decisions can be made on the health of the geared system. Understanding the strengths, weaknesses and constraints of each measurement technology, then capitalizing on these strengths via data fusion, is key to the development of future health monitoring systems.

Appendix A

Design of Experiments

How many experiments are required to verify the data used to build the model reflects the actual process? Looking at the raw sensor data separately, it would be very difficult to apply simple statistical analyses to determine sample size. Once again verifying the complexity of interpreting the data individually. However, if the fused data from the model output is used, simple statistical techniques may be applied to determine the minimum number of samples to verify the objectives of this research. In order to do this several questions must be answered:

1. What are the objectives of this study? The objective of this analysis is to determine if data from 2 accelerometers and one oil debris monitor can be fused into one measurement to predict gear pitting damage.
2. What hypotheses are going to be entertained? The hypothesis: Does the **fused** data from an undamaged gear look significantly different than the **fused** data of a damaged gear?
3. What measurements will be used to address the objectives/hypothesis? The dependent variables are vibration algorithms NA4, NB4, and the oil debris mass. The controlled independent variables are load, rpm, and spur gears.
4. What test statistic should be used to determine if 2 or more populations of Y values are significantly different from one another. In other words, is the fused data from damaged gears different than fused data from undamaged gears. A pooled t-test was chosen.

A pooled t test is a procedure that assumes both populations have the same variance and standard deviation. The name pooled is used because the standard deviation of the 2 samples is pooled to get an estimate of the common standard deviation (Ryan and Joiner (1994)). This test statistic is used to make inferences about 2 means that are independent and the samples are small. When using this statistic, it is assumed that the 2 samples are independent, the samples are randomly selected from normally distributed populations, and at least 1 of the sample sizes is less than 30 and both samples have the same standard deviation. The t-test statistic is calculated as follows (Ryan and Joiner (1994)):

$$t = \frac{\left(\overline{x_1} - \overline{x_2}\right)}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}} \tag{A.1}$$

In order to calculate sample size, it is modified, and trial and error is used to determine sample size.

$$[t(1-\frac{\alpha}{2};\ N_1 + N_2 - 2) + t(1\text{-}\beta;\ N_1 + N_2 - 2)]\sqrt{\frac{1}{N_1}+\frac{1}{N_2}} = \frac{\delta}{S_p} \tag{A.2}$$

The parameters are defined below:

α = type I error rate, reject null hypothesis when null hypothesis is true = only $\alpha \times 100$ percent claim a significant difference in means when there is not. False alarm rates = < .01 (1 percent) = .01
β = type II error rate, fail to reject null hypothesis when null hypothesis is false = only $\beta \times 100$ percent of the time claim there is not a discernable difference in means when their truly is. Damage, but do not indicate damage (missed hits). = .01
$\delta = \mu 1$ to $\mu 2$ = difference in means. How large a difference in means need to be detected? Output of fuzzy logic model is 0-1 = 1
S_p = pooled estimate within population standard deviation (from a previous study). Variation expected in final product properties under repeat conditions. Since levels of damage are separated into 3 levels (0-.33, .33-.66, and .66-1.0) = .33

Examples of calculating sample sizes for this research is listed below. The samples size is a minimum of 16 with damage or no damage sample being a minimum of 5.

Given: $\alpha = .01$; $\beta = .01$; $\delta = 1$; $S_p = .33$

- If $N_1 = 8$ and $N_2 = 8$

$$t(1-\frac{\alpha}{2};\ N_1 + N_2 - 2) = t(0.995,14) = 2.977 = a \tag{A.3}$$

$$t(1 - \beta;\ N_1 + N_2 - 2) = t(0.99,14) = 2.624 = b \tag{A.4}$$

$$\sqrt{\frac{1}{N_1}+\frac{1}{N_2}} = .5 = c \tag{A.5}$$

$$[a + b] * c = 2.80$$

$$\frac{\delta}{S_p} = 3.030 \qquad 2.80 < 3.03 \text{ O.K}$$

- If $N_1 = 5$ and $N_2 = 11$

$$t(1-\frac{\alpha}{2};\ N_1 + N_2 - 2) = t(0.995,14) = 2.977 = a$$

$$t(1 - \beta;\ N_1 + N_2 - 2) = t(0.99,14) = 2.624 = b$$

$$\sqrt{\frac{1}{N_1}+\frac{1}{N_2}} = 0.539 = c$$

$$[a + b] * c = 3.02$$

$$\frac{\delta}{S_p} = 3.030 \qquad 3.02 < 3.03 \text{ O.K}$$

One problem with using the t-test statistic is that it assumes the data are normally distributed. If the data are not normally distributed, nonparametric methods must be used.

The Wilcoxon Rank-Sum Test was also applied to this data. The Wilcoxon Rank-Sum Test assumes the 2 independent samples, null hypothesis is that 2 samples come from same distribution, alternative hypothesis 2 samples different in some way and each sample has more than 10 scores (Triola (1995)). The final output of each experiment was used as the samples for damage and no damage data sets and it varied from 0 to 1. Since their were only 9 experiments with no damage, it was assumed damage was indicated for these 2 additional experiments. An α = .01 was used to identify the z test statistic of 2.575. The table below for no damage/damage lists the output of the model. The data was ranked and the equations are also listed below. This test statistic also found the data sets were significantly different.

TABLE A.1—Experiments with damage/no damage at test completion

No damage	Damage
0.5 (2) 11	1 (6) 20
0.1 (1) 4.5	0.9 (3) 14
0.1 (1) 4.5	0.98 (5) 17
0.1 (1) 4.5	0.95 (4) 15.5
0.1 (1) 4.5	0.5 (2) 11
0.1 (1) 4.5	0.95 (4) 15.5
0.1 (1) 4.5	0.5 (2) 11
0.1 (1) 4.5	1 (6) 20
0.1 (1) 4.5	1 (6) 20
1 (6) 20	0.5 (2) 11
1 (6) 20	0.5 (2) 11
$n_{1=11}$ R_1= 87	$n_{2=11}$ R_2= 166

Rank listed in Table
1-8 (1)'s = (1+2+3+4+5+6+7+8)/8=4.5
9-13 (2)'s =(9+10+11+13)/5= 11
14 (3)
15-16 (4)'s = (15+16)/2=15.5
17 (5)
18-22 (6)'2 = (18+19+20+21+22)/5 = 20

$$\mu_R = \frac{n_1(n_1 + n_2 + 1)}{2} = 126.5$$

$$\sigma_R = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} = 15.23$$

$$z = \frac{R - \mu_R}{\sigma_R} = 2.59$$

Appendix B

Statistical Distributions of Wear Debris

One technique in the literature discussed procedures to detect wear conditions in gear systems by applying statistical distribution methods to particles collected from lubrication systems (Roylance (1989)). The technique involved calculating the mean particle size, variance, kurtosis, and relative kurtosis for debris generated from the gear systems and collected off-line. The wear activity was determined by the calculated size distribution characteristics. In order to apply this data set to on-line oil debris monitor data, the calculations were made for each reading number for each bin using an average particle size calculated for each bin size range. The mean particle size was calculated as:

$$E(d) = \sum_{j=1}^{N} d_j P[d_j] \tag{B.1}$$

where

d_j = average bin size
j = number of bins
$P[d_j]$ = number of particles per average bin size per reading/total number of particles per reading

Then variance kurtosis and relative variance were calculated as follows (Roylance (1989)):

$$Variance = \sum_{j=1}^{N} \left(d_j - E(d)\right)^2 P[d_j] \tag{B.2}$$

$$Kurtosis = \sum_{j=1}^{N} \left(d_j - E(d)\right)^4 P[d_j] \tag{B.3}$$

$$Skewness = \sum_{j=1}^{N} \left(d_j - E(d)\right)^3 P[d_j] \tag{B.4}$$

$$Relative\ Kurtosis = \frac{Kurtosis}{(Variance)^2} \tag{B.5}$$

$$Relative\ Skewness = \frac{Skewness}{(Variance)^{3/2}} \tag{B.6}$$

Using the particle size distribution of the debris, relative kurtosis, mean particle size, and relative kurtosis were calculated for each reading for the experiments with damage. Examples of this data are shown in Figures B.1 and B.2. for experiments 2 and 3. The number of particles per each bin number at test completion, and the particle distribution per each bin for each reading is also shown. Refer to Table 7.1 for the particle size range for each bin. A consistent feature using these statistical parameters could not be found. Setting threshold limits using relative kurtosis, relative skewness or mean particle size would result in a high level of false alarms since it varied significantly for each experiment. This information may be useful for future work in defining different types of failure mechanisms using different particle sizes.

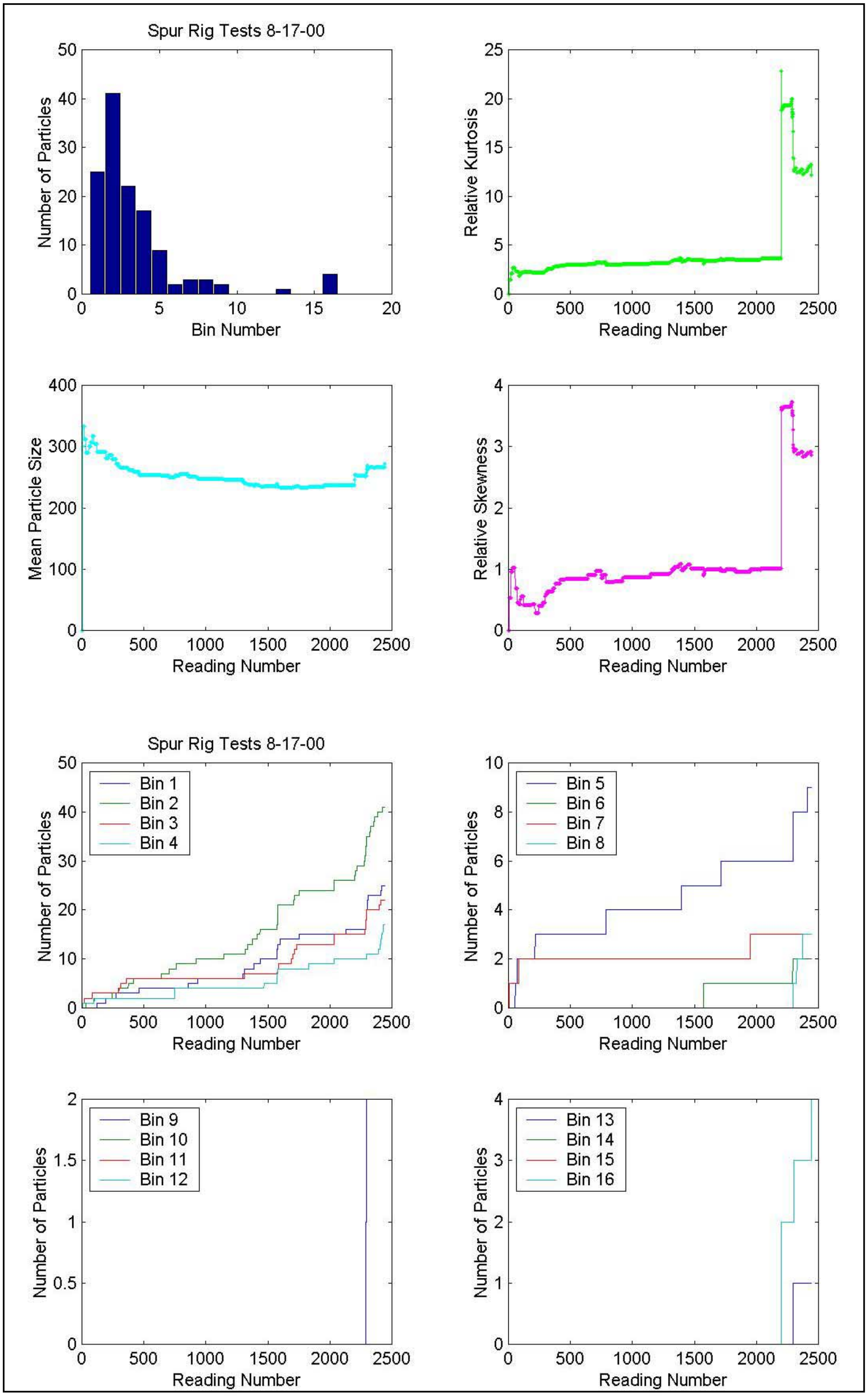


Figure B.1: Statistical distribution methods applied to experiment 2 oil debris data.

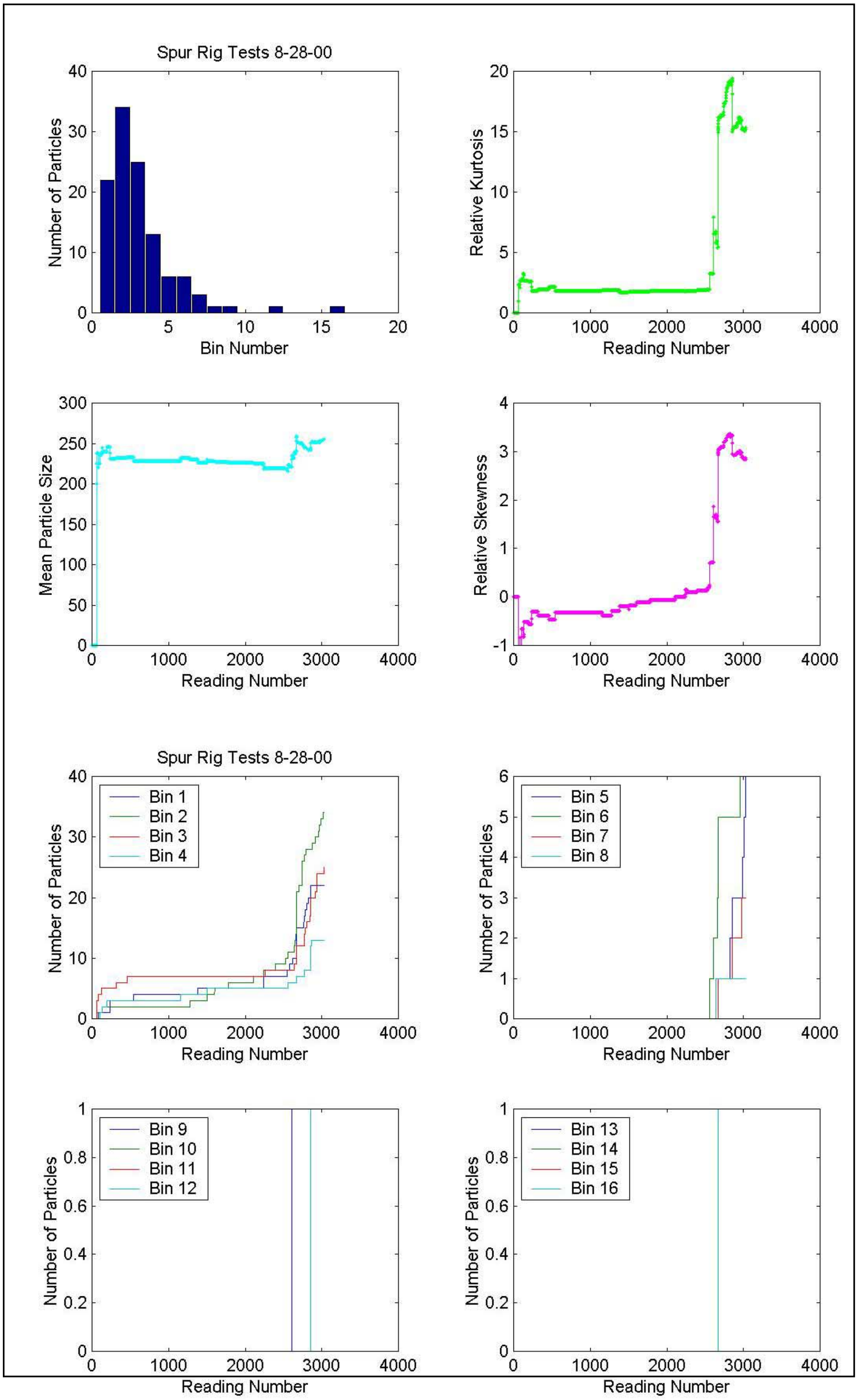


Figure B.2: Statistical distribution methods applied to experiment 3 oil debris data.

Appendix C

Modal Analysis for Selecting Accelerometer Locations

A sensor location check was performed on the housing of the Spur Gear Fatigue Rig gearbox. An accelerometer was mounted at 5 different locations on the housing and an instrumented modal hammer was used to apply a force on the test gear shaft in the direction the meshing forces act on the shaft. The location of the accelerometers in inches is shown in Figure C.1. The input amplitude was measured using a spectrum analyzer and plotted. The plot of the input amplitude indicates the frequencies that are modally active at each location. The coherence function was also calculated and plotted. Coherence is the measure of the amount of output signal that is related to the input signal at a given frequency. A coherence of 1.0 indicates the output is directly related to the input signal at the specified frequency.

Figures C.3 to C.7 show the results of the analysis of the mounting locations on the gearbox housing. Measurements were also plotted in Figure C.2 for the existing sensor on the gear shaft bearing support (measurements are in inches). Combining coherence with minimal modal activity, Location E and B seem to be the best accelerometer location. Since previous tests were performed at location E, will continue to use location E for the accelerometer housing location. Figure C.8 shows the location of the accelerometers on the Spur Gear Fatigue Rig.

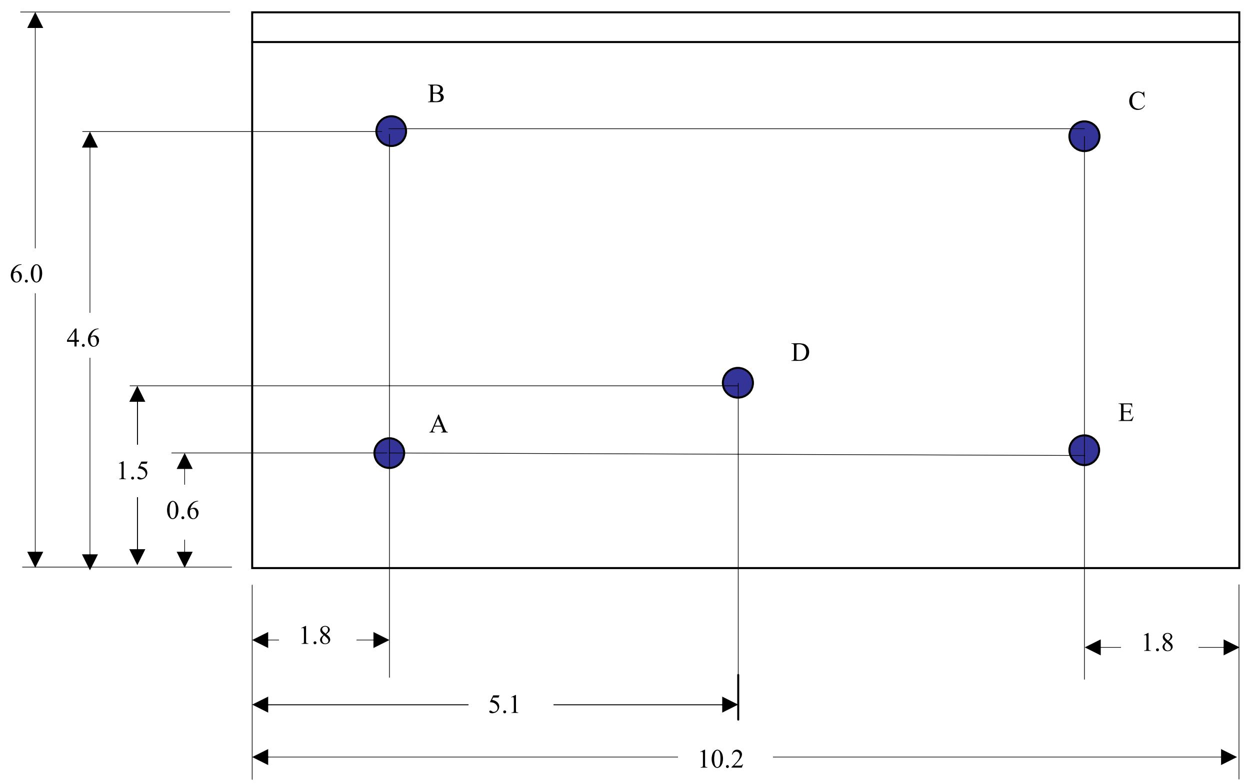


Figure C.1: Gear Box Housing Top View.

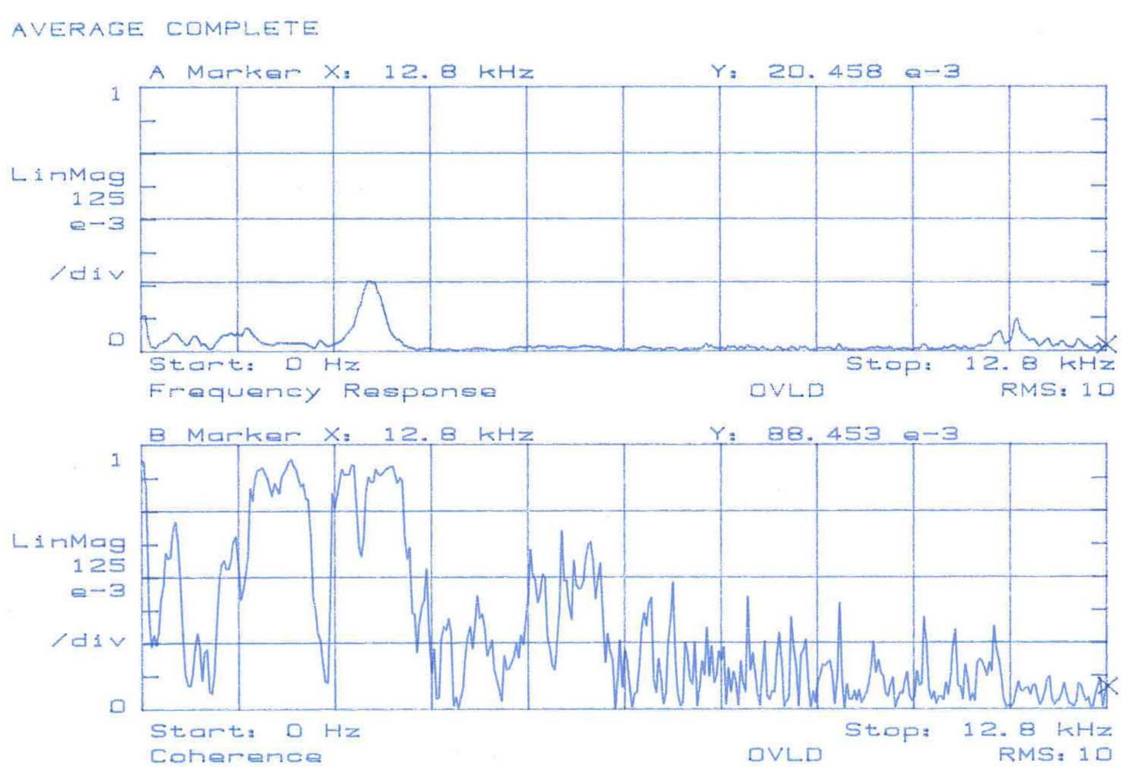


Figure C.2: Accelerometer on Shaft Bearing Support

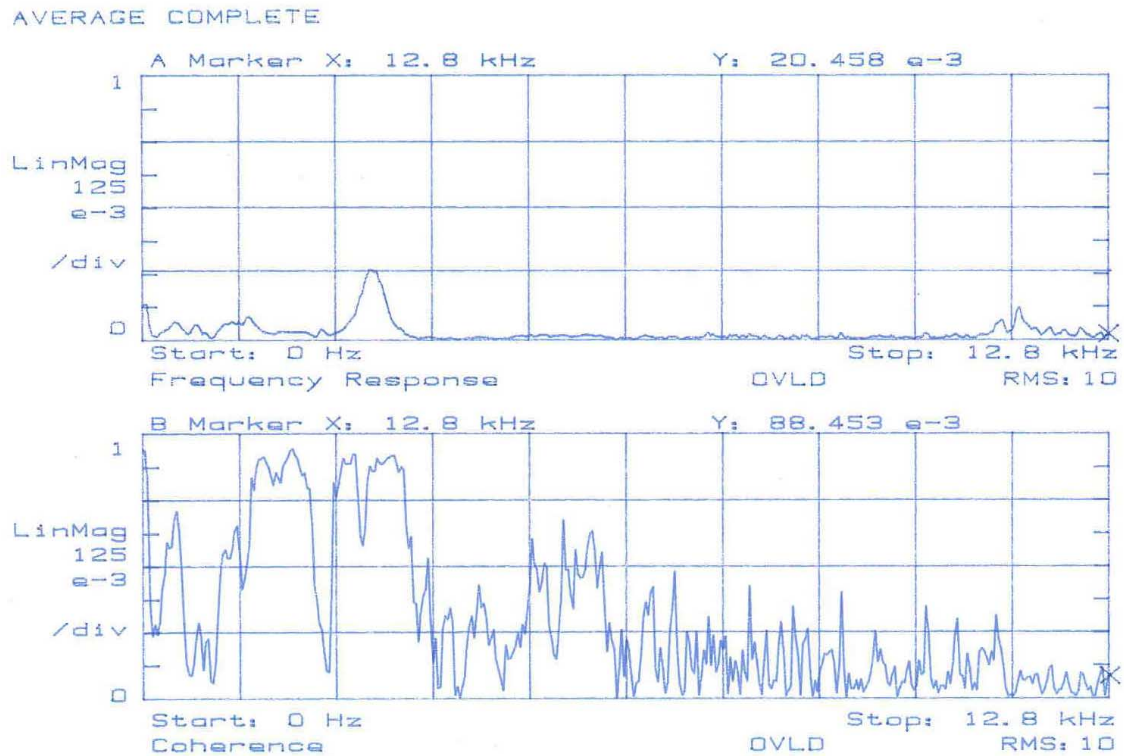


Figure C.3: Accelerometer on Housing Location E.

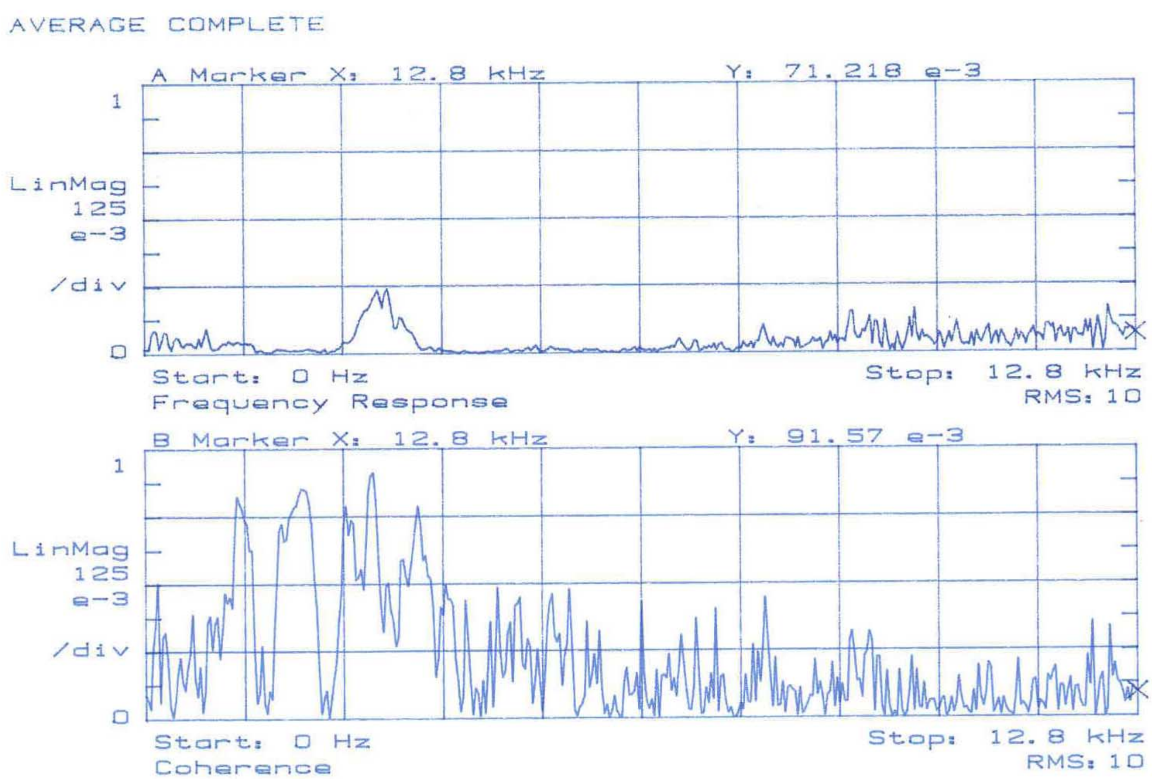


Figure C.4: Accelerometer on Housing Location A.

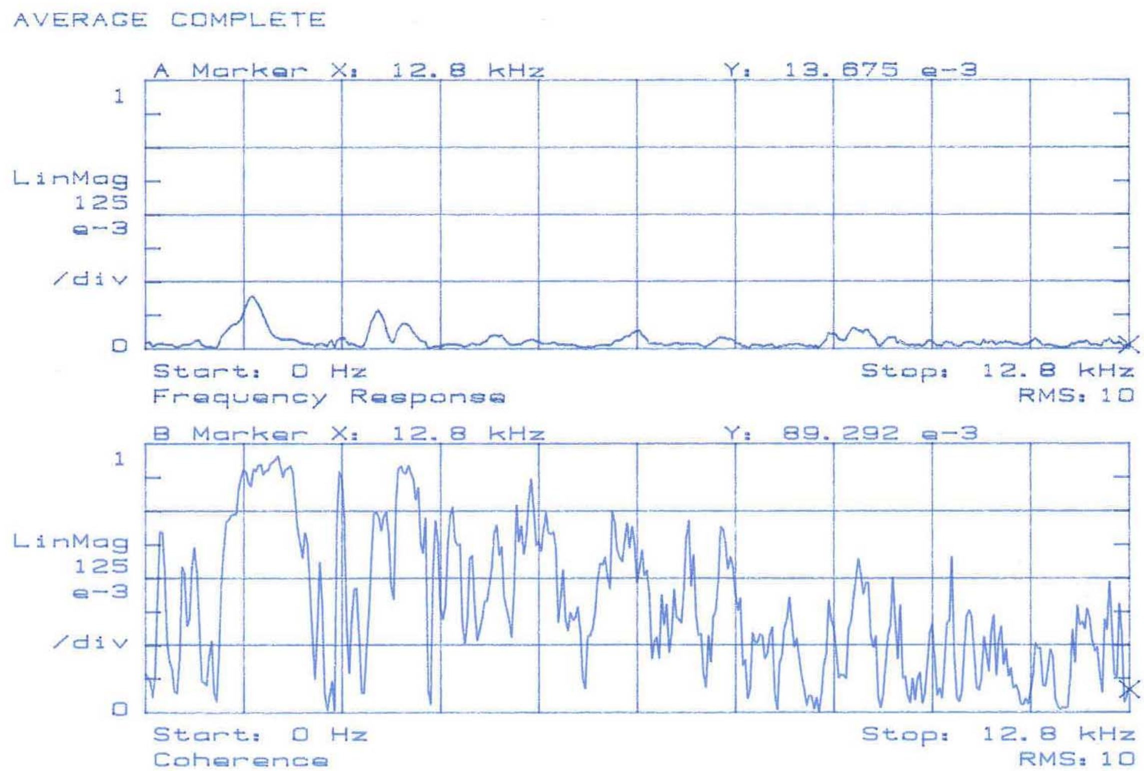


Figure C.5: Accelerometer on Housing Location B.

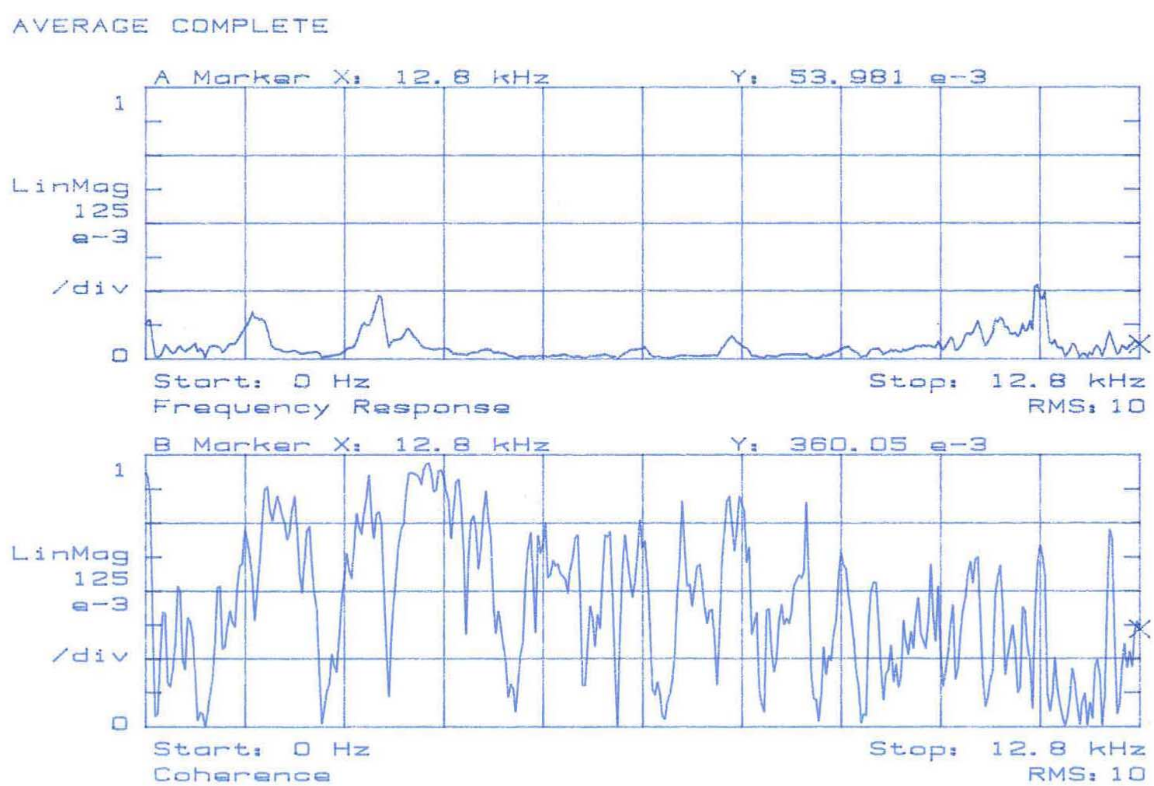


Figure C.6: Accelerometer on Housing Location C.

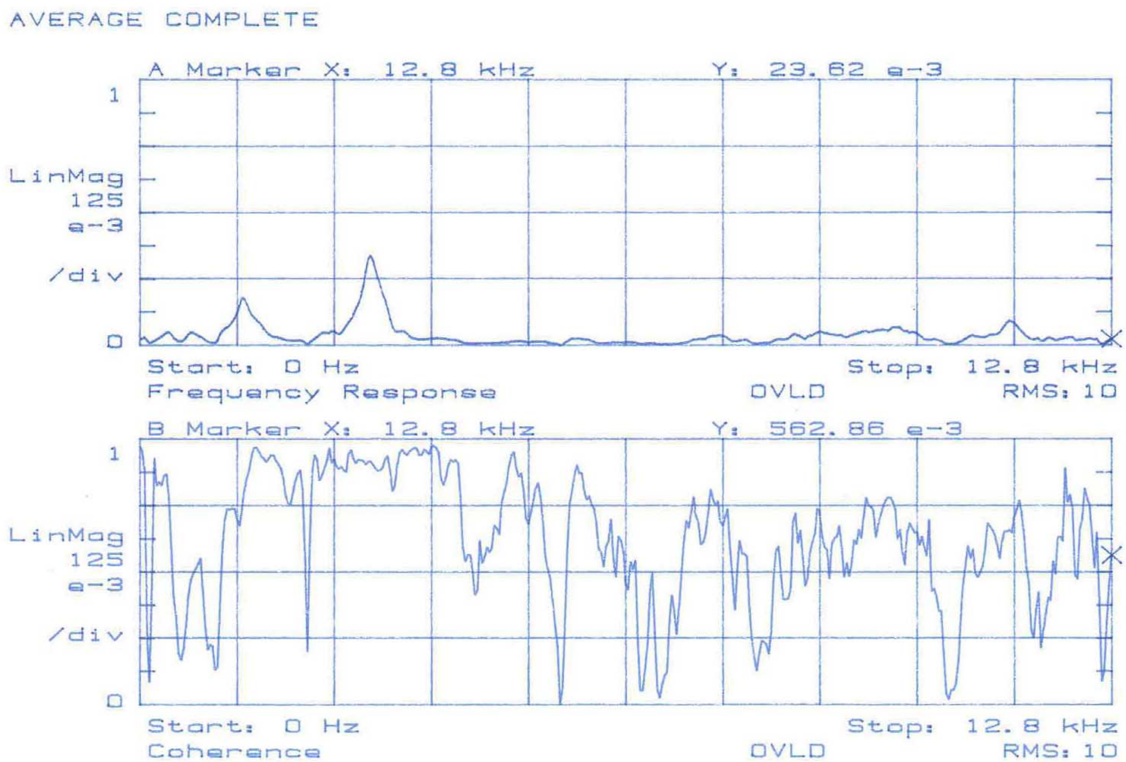


Figure C.7: Accelerometer on Housing Location D.

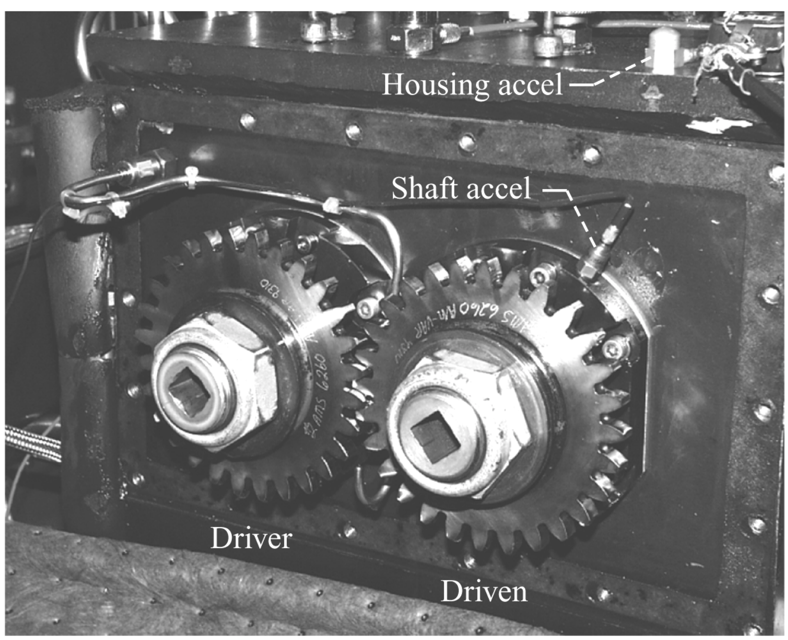


Figure C.8: Accelerometer Locations on Spur Gear Fatigue Rig.

Appendix D

Bayesian Statistics

Many analysis techniques can be used for performing data fusion, and there are no rules regarding what specific fusion technique will work best for a specific application. Decision level fusion processes identity declarations from multiple sensors to achieve a joint declaration of identity. Each sensor performs an identity declaration followed by a process that fuses them to joint multisensor identity declaration. Fuzzy logic was chosen in this analysis for the fusion process. Bayesian inference is another analysis technique that can be used for decision level fusion. Bayesian inference updates the likelihood of a hypothesis given a previous likelihood estimate and additional evidence (observations). The technique may be based on either classical probabilities or subjective probabilities (Hall (1992)).

Bayesian inference can be used to determine the probability that a diagnosis of gear damage is correct given a priori information. The equation for Bayesian inference is:

$$P\left(f_1 \setminus O_n\right) = \frac{P\left(O_n \setminus f_1\right) \cdot P\left(f_1\right)}{\sum_{j=1}^{n} P\left(O_n \setminus f_1\right) \cdot P\left(f_1\right)} \tag{D.1}$$

where $P(f_1 \setminus O_n)$ equals the probability of fault *(f)* given diagnostic output *(O)*, $P(O_n \setminus f_1)$ equals the probability that a diagnostic output *(O)* is associated with fault *(f)*, and $P(f_1)$ is the probability of *(f)* occurring (Kacprzynski (2001)).

Bayesian inference was not chosen for the fusion process for several reasons. The first was defining a prior likelihood's for the vibration and oil debris features. Bayesian inference requires knowledge about the diagnostic system to generate the a priori distributions. Integration of the vibration and oil debris features is a new diagnostic technique. The data integrating two vibration algorithms and the oil debris sensor does not exist outside the scope of this thesis. Bayesian inference requires a priori probabilities of the hypotheses that did not exist outside the scope of this research. Limited knowledge and data made it impossible to translate the preliminary data into a probability distributions.

Another reason is that the data did not necessarily follow a normal distribution. Both data sets showed a deviation from normality in terms of kurtosis and skewness. The oil debris data had large amounts of skewness and the vibration algorithms are defined in terms of kurtosis. All parameters showed a deviation from normality in terms of kurtosis and skewness. A normal distribution is symmetric and has no kurtosis (how heavy the tails of distribution are). Since the data does not follow a normal distribution is was not clear how this would affect the resulting posterior distribution (Iverson (1984)).

The complexity of the data due to multiple hypothesis and multiple conditional dependent events made it difficult to define levels of probability for each scenario. For example the different levels of damage indicated by each sensor and the resulting different states of the system. The complexity of the time dependent data made it difficult to define simple probabilities. Assumptions can be made based on a probabilities for each experiment that relate to the gear state based on the final test reading. This does not provide on-line condition maintenance for the duration of the test. Instead, the data is processed through the fuzzy logic model and normalized to 1, and this data is then used as an input to the Bayesian inference system.

The success of the inference system depends on its ability to represent knowledge about the application domain. Bayesian statistics is useful when applied to fault classification, where a large amount of fault data is acquired on different types of faults (Erdley and Hall (1998)). Fuzzy logic was a better choice for this application by establishing relationships between different measurement technologies to identify the state of the gear with limited data.

REFERENCES

Astridge, Derek G. (1987): HUM—Health and Usage Monitoring of Helicopter Mechanical Systems. Condition Monitoring ′87: Proceedings of an International Conference on Condition Monitoring, pp. 143−162.

Astridge, D.G. (1989): Helicopter Transmissions—Design for Safety and Reliability. Proc. Inst. Mech. Engrs., vol. 203, pp. 123−138.

Aviation Safety and Security Program, the Helicopter Accident Analysis Team: Final Report of the Helicopter Accident Analysis Team (1998).

Becker, Kimberly C.; Byington, C.S.; Forbes, N.A.; and Nickerson, G.W. (1998): Predicting and Preventing Machine Failures. The Industrial Physicist, Vol. 4, No. 4, pp. 20–23.

Bowman, Christopher (2001): Data Fusion Short Course. Data Fusion and Neural Networks. Broomfield, Colorado.

Byington, C.S.; Merdes, T.A.; and Kozlowski, J.D. (1999): Fusion Techniques for Vibration and Oil Debris/Quality in Gearbox Failure Testing. Proceedings of the Condition Monitoring International Conference, M.H. Jones and D.G. Sleeman, eds., Coxmoor Publishing, Oxford, England, pp. 113–128.

Campbell, R.L.; Byington, C.S.; and Lebold, M.S. (2000): Generation of HUMS Diagnostic Estimates Using Transitional Data. Proceedings of 13th International Congress on Condition Monitoring and Diagnostic Engineering Management, Henry C. Pusey and Raj B.K.N. Rao, eds., The Society for Machinery Failure Prevention Technology, Haymarket, VA, pp. 587–595.

Choy, F.K.; Huang, S.; Zakrajsek, J.J; Handschuh, R.F.; and Townsend, D.P. (1994): Vibration Signature Analysis of a Faulted Gear Transmission System. NASA TM–106623.

Dasarathy, Belur V. (2001): Information Fusion—what, where, why when, and how? Information Fusion, vol. 2, issue 2, pp. 75−76.

Dempsey, Paula J. (2000): A Comparison of Vibration and Oil Debris Gear Damage Detection Methods Applied to Pitting Damage. NASA/TM—2000-210371.

Dempsey, Paula J.; and Zakrajsek, James J. (2001): Minimizing Load Effects on NA4 Gear Vibration Diagnostic Parameter. NASA/TM—2001-210671.

Dempsey, Paula J. (2001): Gear Damage Detection Using Oil Debris Analysis. NASA/TM—2001-210936.

Erdley, J.D.; and Hall, D.L. (1998): Improved Fault Detection Using Multisensor Data Fusion. Proceedings of the 52nd Meeting of the Society for Machinery Failure Prevention Technology, pp. 337−346.

Feng, Xu; Yubing, Yang; and Baosheng, Wang (2000): Realization of Automatic On-Line Fault Diagnosis of Rotating Machine by Fuzzy Identification. Proceedings of the 8th International Symposium on Transport Phenomena and Dynamics of Rotating Machinery, vol. II.

Forror, Mark (2000): A Mind of its Own. Rotor & Wing, vol. 34, No. 8, pp. 42–47.

Fuzzy Logic Toolbox for Use With MATLAB® (1998): MathWorks, Inc., Natick, MA.

Garga, Amulya K.; and Hall, David L. (1999a): A Unifying Approach for Rotor-Craft Information Systems. Proceedings of the 55th AHS International Annual Forum, American Helicopter Society, Alexandria, VA, pp. 837−843.

Garga, A.K.; and Hall, D.L. (1999b): Perspectives on the Progress of Data Fusion for Soldiers. Conference Record of the Thirty-Third Asilomar Conference on Signals, Systems, and Computers, vol. 1, pp. 402−406.

Gibson, R.E.; Hall, D.L.; and Stover, J.A. (1994): An Autonomous Fuzzy Logic Architecture for Multisensor Data Fusion. Proceedings of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, pp. 143−150.

Hall, David L. (1992): Mathematical Techniques in Multi-Sensor Data Fusion. Artech House, Boston, MA.

Hall, David L.; and Kasmala, G. (1996): Visual Programming Environment for Multisensor Data Fusion. Proceedings of the SPIE International Society for Optical Engineering Meeting, issue 2764, pp. 181−187.

Hall, David L; and Llinas, J. (1997): An Introduction to Multisensor Data Fusion. Proc. IEEE, vol. 85, no. 1, pp. 6−23.

Hall, David L.; Garga, A.K.; and Stover, J. (1999a): Machinery Fault Classification: The Case for a Hybrid Fuzzy Logic Approach. Society for Machinery Failure Prevention Technology, H.C. Pusey, S.C. Pusey, and W.R. Hobbs, eds., Haymarket, VA, pp. 241−252.

Hall, David L.; Garga, Amulya K. (1999b): Pitfalls in Data Fusion (and How to Avoid Them). International Society of Information Fusion, FUSION'99, Final Program, Sunnyvale, CA, pp. 429−436.

Handschuh, Robert F. (1995): Thermal Behavior of Spiral Bevel Gears. NASA TM−106518.

Handschuh, Robert F. (2001): Testing of Face-Milled Spiral Bevel Gears at High-Speed and Load. NASA/TM—2001-210743.

Hannah, P.; Starr, A.; and Ball, A. (2000): Decisions in Condition Monitoring—An Exemplar for Data Fusion Architecture. Proceedings of the 3rd International Conference on Information Fusion, vol. 1.

Howard, Paul L.; Roylance, B.; Reintjes, J.; and Schultz, A. (1998): New Dimensions in Oil Debris Analysis—The Automated, Real Time, On Line Analysis of Debris Particle Shape. Naval Research Lab, Washington, DC.

Howard, Paul L. and Reintjes, J.F. (1999): A Straw Man for the Integration of Vibration and Oil Debris Technologies. Helicopter Health and Usage Monitoring Systems Workshop, G.G. Forsyth, ed., Defense Science and Technology Organization General Document 197, no. 1, pp. 131−136.

Howe, B.; and Muir, D. (1998): In-Line Oil Debris Monitor (ODM) for Helicopter Gearbox Condition Assessment. AD–a347 503, Defense Technical Information Center, Ft. Belvoir, VA.

Hunt, Trevor M. (1993): Handbook of Wear Debris Analysis and Particle Detection in Liquids. Elsevier Applied Science, London, 1993.

Iversen, Gudmund R. (1984): Bayesian Statistical Inference. Sage Publications, Newbury Park, CA.

Jang, Jyh-Shing Roger; and Sun, Chuen-Tsai (1995): Neuro-Fuzzy Modeling and Control. Proc. IEEE, vol. 83, no. 3, pp. 378–406.

Jantzen, Jan (1999a): Tutorial On Fuzzy Logic. Pub. No. 98–E–868, Technical University of Denmark.

Jantzen, Jan (1999b): Design of Fuzzy Controllers. No. 98–E–864, Technical University of Denmark.

Kacprzynski, Gregory J.; Roemer, M.J.; Orsagh, R.F. (2001): Assessment of Data and Knowledge Fusion Strategies for Diagnostics and Prognostics. Society for Machinery Failure Prevention Technology, H.C. Pusey, S.C. Pusey, and W.R. Hobbs, eds., Haymarket, VA, pp. 341–350.

Kasabov, Nikola K. (1996): Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering. MIT Press, Cambridge, MA.

Kotanchek, M.E. (1995): CBM/IPD Process Model. Pennsylvania State Applied Research Laboratory Technical Memorandum, File No. 95–113.

LabVIEW™ Basics I Course Manual (1998): Course Software Version 4.0, pt. no. 320628E–01.

Land, James E. (1998): HUMS—Benefits Emerge. Helicop. Wld., vol. 17, no. 9, pp. 19–23.

Larder, Brian D.(1994): Vibration Monitoring of Helicopter Transmissions: Lessons Learned and Future Developments. Health and Usage Monitoring of Rotorcraft Transmission Systems—Review of Service Experience, Institution of Mechanical Engineers, London.

Larder, B.D. (1999): Helicopter HUM/FDR: Benefits and Developments. Proceedings of the 55th AHS International Annual Forum, American Helicopter Society, Alexandria, VA.

Learmount, David (2000): Rotary Woes. Flight International, no. 4725, vol. 157, pp. 34–35.

Lewicki, David G. (2001): Gear Crack Propagation Path Studies-Guidelines for Ultra-Safe Design. NASA/TM—2001-211073.

Luo, R.C.; and Su, K.L. (1999): A Review of High-Level Sensor Fusion Approaches and Its Applications. IEEE SICE RSJ International Conference on Multisensor Fusion and Integration for Intelligent Systems, Taipei, Japan.

Lynwander, Peter (1983): Gear Drive Systems: Design and Application. Marcel Dekker, New York, NY.

Mamdani, E.H.; and Assilan, S. (1975): An Experiment in Linguistic Synthesis With a Fuzzy Logic Controller. International Journal of Man-Machine Studies, Vol. 7, No. 1, pp.1–13.

McFadden, P.D. (1987): Examination of a Technique for the Early Detection of Failure in Gears by Signal-Processing of the Time Domain Average of the Meshing Vibration. Mechanical Systems and Signal Processing, vol. 1, no. 2, pp. 173−183.

McGonigal, Daniel Lewis (1997): A Comparison of Automated Reasoning Techniques for Condition-Based Maintenance. M.A. Thesis, Pennsylvania State Univ.

MetalSCAN™ User's Manual. Setting Alarm Limits, C000833, Revision 0, GasTOPS, Ontario, Canada.

MetalSCAN™ User's Manual. C000868, rev. 2, GasTOPS, Ontario, Canada.

Miller, J.L.; and Kitaljevich, D. (2000): In-Line Oil Debris Monitor for Aircraft Engine Condition Assessment. Aerospace Conference Proceedings, 2000 IEEE, vol. 6, pp. 49−56.

Orsagh, Rolf F., Savage, C.J., and McClintic, K. (2000): Development of Metrics for Mechanical Diagnostic Technique Qualification and Validation. Proceedings of 13th International Congress on Condition Monitoring and Diagnostic Engineering Management, Henry C. Pusey and Raj B.K.N. Rao, eds., The Society for Machinery Failure Prevention Technology, Haymarket, VA, pp. 515−526.

Pouradier, Jean-Marc; and Trouvé, Michel (2001): An Assessment of Eurocopter Experience in HUMS Development and Support. Proceedings of the 57th AHS International Annual Forum, American Helicopter Society, Alexandria, VA.

Roylance, B.J. (1989): Monitoring Gear Wear Using Debris Analysis—Prospects for Establishing a Prognostic Method. Proceedings of the 5th International Congress on Tribology, Kenneth Holmberg and Ilkka Nieminen, eds., vol. 4, p. 85.

Roylance, B.J. (1997): Some Current Developments for Monitoring the Health of Military Aircraft Using Wear Debris Analysis Techniques. ASME—Trib., R.S. Cowan, ed., vol. 7, pp. 55−60.

Ryan, Barbara F.; and Joiner, Brian L. (1994): Minitab Handbook. Duxbury Press, Belmont, CA.

Slemp, Mark; and Skeirik, Robert (1999): Using Statistics to Avoid Vibration False Alarms. P/PM Technology.

Steinberg, Alan N.; Bowman, C.L; White, F.E. (1999): Revisions to the JDL data fusion model. SPIE Conference on Sensor Fusion: Architectures, Algorithms, and Applications III.

Stewart, R.M. (1977): Some Useful Data Analysis Techniques for Gearbox Diagnostics. Machine Health Monitoring Group Report MHM/R/10/77, University of Southampton.

Stewart, Ronald M.; and Ephraim, P. (1997): Advanced HUMS and Vehicle Management Systems Implemented Through an IMA Architecture. Annual Forum Proceedings of the American Helicopter Society, vol. 53/V2, pp. 1257−1266.

Townsend, Dennis P. (1991): Dudley's Gear Handbook. McGraw-Hill, New York, NY.

Townsend, Dennis P. (1997): Gear and Transmission Research at NASA Lewis Research Center, NASA TM−107428.

Triola, Mario F. (1995): Elementary Statistics. Addison-Wesley, Reading, MA.

Tustin Technical Institute (1996): Digital Signal Processing and Data Analysis Course.

Waltz, Edward L. (1986): Data Fusion for C^3I Systems. CfI Handbook: Command, Control, Communications, Intelligence, 1st ed. EW Communications, Palo Alto, CA, pp. 217−226.

Yeh, Zong-Mu (1999): A Systematic Method for Design of Multivariable Fuzzy Logic Controls Systems. IEEE Trans. Fuzzy Sys., vol. 7, no. 6, pp. 741–752.

Zadeh, Lotfi A. (1973): Outline of a New Approach to the Analysis of Complex Systems and Decision Processes. IEEE Trans. Syst. Man Cybern., vol. SMC–3, no. 1, pp. 28–44.

Zadeh, Lotfi Asker (1992): Fuzzy Logic Advanced Concepts and Structures. IEEE Educational Activities Board. IEEE, Piscataway, NJ.

Zakrajsek, James J. (1989): An Investigation of Gear Mesh Failure Prediction Techniques. NASA TM–102340.

Zakrajsek, James J.; Townsend, D.P.; Oswald, F.B.; and Decker, H.J. (1992): Analysis and Modification of a Single-Mesh Gear Fatigue Rig for Use in Diagnostic Studies. NASA TM–105416.

Zakrajsek, J.J.; Townsend, D.P.; and Decker, H.J. (1993): An Analysis of Gear Fault Detection Methods as Applied to Pitting Fatigue Failure Data. NASA TM–105950.

Zakrajsek, James J.; Handschuh, R.F.; and Decker, H.J. (1994a): Application of Fault Detection Techniques to Spiral Bevel Gear Fatigue Data. NASA TM–106467.

Zakrajsek, James J.; Decker, H. J.; Handschuh, R.F. (1994b): An Enhancement to the NA4 Gear Vibration Diagnostic Parameter.

Zakrajsek, James J. (1994): A Review of Transmission Diagnostics Research at NASA Lewis Research Center, NASA TM–106746.

Zakrajsek, James J.; Decker, H.J.; Handschuh, R.F.; and Lewicki, D.G. (1995a): Detecting Gear Tooth Fracture in a High Contact Ratio Face Gear Mesh. NASA TM–106822.

Zakrajsek, J.J.; Townsend, D.P.; Lewicki, D.G.; Decker, H.J., and Handschuh, R.F. (1995b): Transmission Diagnostic Research at NASA Lewis Research Center, NASA TM–106901.

Zakrajsek, James J.; and Lewicki, David G. (1996): Detecting Gear Tooth Fatigue Cracks in Advance of Complete Fracture, NASA TM–107145.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)

2. REPORT DATE
March 2003

3. REPORT TYPE AND DATES COVERED
Technical Memorandum

4. TITLE AND SUBTITLE
Integrating Oil Debris and Vibration Measurements for Intelligent Machine Health Monitoring

5. FUNDING NUMBERS
WU–712–30–13

6. AUTHOR(S)
Paula J. Dempsey

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
National Aeronautics and Space Administration
John H. Glenn Research Center at Lewis Field
Cleveland, Ohio 44135–3191

8. PERFORMING ORGANIZATION REPORT NUMBER
E–13109

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
National Aeronautics and Space Administration
Washington, DC 20546–0001

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
NASA TM—2003-211307

11. SUPPLEMENTARY NOTES
This report was submitted as a dissertation in partial fulfillment of the requirements for the degree Doctor of Philosophy in Engineering Science to the University of Toledo, Toledo, Ohio, May 2002. Responsible person, Paula J. Dempsey, organization code 5950, 216–433–3398.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Unclassified - Unlimited
Subject Category: 01 Distribution: Nonstandard

Available electronically at http://gltrs.grc.nasa.gov
This publication is available from the NASA Center for AeroSpace Information, 301–621–0390.

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 words)***

A diagnostic tool for detecting damage to gears was developed. Two different measurement technologies, oil debris analysis and vibration were integrated into a health monitoring system for detecting surface fatigue pitting damage on gears. This integrated system showed improved detection and decision-making capabilities as compared to using individual measurement technologies. This diagnostic tool was developed and evaluated experimentally by collecting vibration and oil debris data from fatigue tests performed in the NASA Glenn Spur Gear Fatigue Rig. An oil debris sensor and the two vibration algorithms were adapted as the diagnostic tools. An inductance type oil debris sensor was selected for the oil analysis measurement technology. Gear damage data for this type of sensor was limited to data collected in the NASA Glenn test rigs. For this reason, this analysis included development of a parameter for detecting gear pitting damage using this type of sensor. The vibration data was used to calculate two previously available gear vibration diagnostic algorithms. The two vibration algorithms were selected based on their maturity and published success in detecting damage to gears. Oil debris and vibration features were then developed using fuzzy logic analysis techniques, then input into a multi sensor data fusion process. Results show combining the vibration and oil debris measurement technologies improves the detection of pitting damage on spur gears. As a result of this research, this new diagnostic tool has significantly improved detection of gear damage in the NASA Glenn Spur Gear Fatigue Rigs. This research also resulted in several other findings that will improve the development of future health monitoring systems. Oil debris analysis was found to be more reliable than vibration analysis for detecting pitting fatigue failure of gears and is capable of indicating damage progression. Also, some vibration algorithms are as sensitive to operational effects as they are to damage. Another finding was that clear threshold limits must be established for diagnostic tools. Based on additional experimental data obtained from the NASA Glenn Spiral Bevel Gear Fatigue Rig, the methodology developed in this study can be successfully implemented on other geared systems.

14. SUBJECT TERMS
Vibration; Debris; Wear; Gear teeth; Pitting; Systems health monitoring; Fuzzy systems; Multisensor fusion

15. NUMBER OF PAGES
122

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	